PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12Q 1/68, C07H 15/12
A01K 45/00

A1

(11) International Publication Number: **WO 92/13102**

(43) International Publication Date: 6 August 1992 (06.08.92)

(21) International Application Number: PCT/US92/00340

(22) International Filing Date: 15 January 1992 (15.01.92)

(30) Priority data:
642,342 15 January 1991 (15.01.91) US

(71) Applicant *(for all designated States except US)*: GENMARK [US/US]; 421 Wakara Way, Salt Lake City, UT 84108 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)* : Georges, Michel [BE/US]; 4880 South 2870 East, Salt Lake City, UT 84117 (US). MASSEY, Joseph, M. [US/US]; 4297 South Zarahemia Drive, Salt Lake City, UT 84124 (US).

(74) Agents: IHNEN, Jeffrey, L. et al.; Venable, Baetjer, Howard & Civiletti, Suite 1000, 1201 New York Avenue, N.W., Washington, DC 20005 (US).

(81) Designated States: AT (European patent), AU, BB, BE (European patent), BF (OAPI patent), BG, BJ (OAPI patent), BR, CA, CF (OAPI patent), CG (OAPI patent), CH (European patent), CI (OAPI patent), CM (OAPI patent), CS, DE (European patent), DK (European patent), ES (European patent), FI, FR (European patent), GA (OAPI patent), GB (European patent), GN (OAPI patent), GR (European patent), HU, IT (European patent), JP, KP, KR, LK, LU (European patent), MC (European patent), MG, ML (OAPI patent), MN, MR (OAPI patent), MW, NL (European patent), NO, PL, RO, RU, SD, SE (European patent), SN (OAPI patent), TD (OAPI patent), TG (OAPI patent), US.

Published
With international search report.

(54) Title: POLYMORPHIC DNA MARKERS IN BOVIDAE

(57) Abstract

Described is a collection of bovine genomic clones that map to polymorphic loci in bovids. Said clones will find utility in genetic identification, gene mapping and selective breeding.

Codes use
applications und

AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Fas
BG	Bulgaria
BJ	Benin
BR	Brazil
CA	Canada
CF	Central Afri
CG	Congo
CH	Switzerland
CI	Côte d'Ivoir
CM	Cameroon
CS	Czechoslova
DE	Germany
DK	Denmark
ES	Spain

FOR THE PURPOSES OF INFORMATION ONLY

l to identify States party to the PCT on the front pages of pamphlets publishing international
r the PCT.

FI	Finland	ML	Mali
FR	France	MN	Mongolia
GA	Gabon	MR	Mauritania
GB	United Kingdom	MW	Malawi
GN	Guinea	NL	Netherlands
GR	Greece	NO	Norway
HU	Hungary	PL	Poland
IE	Ireland	RO	Romania
IT	Italy	RU	Russian Federation
JP	Japan	SD	Sudan
KP	Democratic People's Republic of Korea	SE	Sweden
		SN	Senegal
KR	Republic of Korea	SU	Soviet Union
LI	Liechtenstein	TD	Chad
LK	Sri Lanka	TG	Togo
LU	Luxembourg	US	United States of America
MC	Monaco		
MG	Madagascar		

TITLE OF THE INVENTION

POLYMORPHIC DNA MARKERS IN BOVIDAE

CROSS-REFERENCE TO RELATED APPLICATIONS

The present invention is a continuation-in-part of U.S. Ser. No. 642,342, filed January 15, 1991, incorporated herein by reference.

FIELD OF THE INVENTION

The invention relates to gene mapping, selective breeding and genetic identification in domestic animals.

BACKGROUND OF THE INVENTION

The publications and other materials used herein to illuminate the background of the invention, and in particular, cases to provide additional details respecting the practice, are incorporated by reference and for convenience are numerically referenced in the following text and respectively grouped in the appended bibliography.

Until recently, artificial selection has relied on the biometrical evaluation of individual breeding values from an animal's own performance and from performance of its relatives (136). This biometrical strategy is based on relatively simple genetic premises, operating within a "black box". Briefly, the majority of economically important traits are so-called complex or quantitative traits, meaning that the phenotype of an animal is determined by both environment and a large number of genes with individually small, additive effects. The proportion of the phenotypic variation observed in a given population that is genetic in nature is the heritability of the trait. Substantial genetic progress has been obtained using this approach. One of the

powers of this biometrical approach is that it obviates the need for any detailed molecular knowledge of the underlying genes or Economic Trait Loci.

However, it is believed that the molecular identification of these Economic Trait Loci could increase genetic response by affecting both time and accuracy of selection, through a procedure called Marker Assisted Selection (91, 96). One strategy towards the isolation of Economic Trait Loci relies on the use of DNA Sequence Polymorphisms as genetic markers in linkage studies. This approach, paradoxically referred to as "Reverse Genetics" (138), will be described in detail in this introduction. Moreover, we propose a new concept called "Velogenetics", or the combined use of Marker Assisted Introgression and germ-line manipulations to shorten the generation interval of domestic species (especially cattle), which will allow the rapid and efficient introgression of mapped Economic Trait Loci between genetic backgrounds.

I. DNA SEQUENCE POLYMORPHISM (DSP)

A. Types of DNA Sequence Polymorphism

The typical mammalian genome is composed of an approximately $3x10^9$ base pairs long DNA stretch, divided over a species-specific number of chromosomes, and containing all the information required for the proper development and functioning of a normal being. Each individual has two copies of this message: one paternal in origin and one maternal. Although their overall architecture and content are virtually identical, the paternal and maternal DNA sequences exhibit subtle "allelic" differences, hereinafter referred to as DNA Sequence Polymorphisms or "DSP". The DSP that can be recognized in a given population are the molecular basis of the genetic component of the observed phenotypic variance. One can distinguish three types of DSP.

1. Single Base Pair Polymorphisms

As their name implies, these DSP are due to single base pair differences distinguishing alleles. These can be either base pair substitutions - transitions (Purine to Purine or Pyrimidine to Pyrimidine) and transversions (Purine to Pyrimidine and vice versa) -, or the insertion/deletion of a single base pair.

The frequency of single base pair polymorphism is measured by the nucleotide diversity, π, or average heterozygosity per nucleotide site (1). The nucleotide diversity has been estimated from Restriction Fragment Length Polymorphisms at 0.002 for human (2), and at 0.0007 in cattle (3,4). This means that on the average a human will be heterozygous for one every 500 nucleotides, and a cow for one every 1,500 nucleotides.

One type of single base pair polymorphism deserves special attention: the CpG to TpG transition. The cytosine in the CpG dinucleotide sequence is known to be the substrate of an eucaryotic methylase, which will add a methyl group in position 5 of the pyrimidine ring, if the cytosine of the complementary CpG dinucleotide is itself methylated. Deamination of a 5-methylcytosine generates a thymine, blurring the task of the DNA repair machinery which will half of the time resolve the ensuing mismatch by replacing the original guanine instead of the mutated thymine. As a consequence, cytosines in the CpG doublet exhibit mutation rates at least ten times higher than other nucleotides, and hence are rich sources of single base pair polymorphisms (4, 5).

2. DNA Sequence Rearrangements

In this kind of DSP, the difference between allelic variants involves DNA sequence rearrangements such as the insertion or deletion of a stretch of DNA, DNA sequence inversions and duplications.

Although there is a wide spectrum of molecular mechanisms susceptible to generate such chromosomal re-

arrangements, it is well established that mobile genetic elements significantly contribute to this kind of DSP.

In lower eukaryotes such as Drosophila and yeast, rearrangements involving transposable elements account for a large proportion of new mutations detected in these organisms (6). In the mouse, retrovirus-like sequences or retrotransposons have been shown to act as insertional mutagens (7-11), and different strains of mice exhibit substantial heterogeneity with respect to the numbers and chromosomal sites of endogenous proviruses (12). Variation in the distribution of endogenous retroviruses has been demonstrated in poultry as well.

In the human, at least 10% of the genome is known to be composed of retroposon-like sequences. Evidence for a role of these sequences in human genetic variability and disease stems from several reports of de novo mutations due to these sequences: a mutation in the human Low Density Lipoprotein receptor gene giving rise to familial hypercholesterolemia is caused by a deletion brought about by an intrastrand recombination event between two Alu sequences (13); L1 insertions were found to inactivate the factor VIII gene in hemophilia A patients (14); a c-myc rearrangement in a breast carcinoma was found to be due to insertion of an L1 element (15); an Alu transposition event has been documented in human lung carcinoma cells (16); and an homologous recombination between the LTRs of a human retrovirus-like element was shown to cause a 5 Kb deletion polymorphism. Recently, Wong et al. (17) reported evidence of human DNA polymorphism arising through DNA-mediated, rather than RNA-mediated, transfer between autosomes as well.

3. Expansion-Contraction Type Polymorphism

A significant proportion of the eucaryotic genome is composed of sequences widely termed "satellite DNA,"

sharing a common organization: a sequence motif, varying in length between one and several thousand nucleotides, repeated in a head-to-tail or so-called tandem arrangement. Depending on the methodology originally used for their study, i.e. isopycnic centrifugation, pulsed field gel electrophoresis, agarose gel electrophoresis or polyacrylamide gel electrophoresis, satellite sequences were grouped into four size classes: macro-, midi-, mini- and micro-satellites. Minisatellites are also known as Variable Number of Tandem Repeats (VNTRs) (18-21). While macro-satellites seem to be confined to heterochromatic regions (22), mini- and micro-satellites have been found scattered throughout the genome with, however, clustering of mini-satellites (23-34). In the human, minisatellite clusters seems to be particularly abundant in proterminal regions (35). The only midi-satellite described as such today, has been mapped to the short arm of chromosome 1 (36). In the human, the polydeoxyadenylate tract of Alu repetitive elements are also caracterized by length variation and are thus an abundant source of genetic markers as well (37).

The function, if any, of satellite sequences, whether macro-, midi-, mini- or micro-, is essentially unknown. An important feature of all satellite sequences is that the maintenance of their tandemly repeated organization is dependent on the concerted evolution of the repeats. This concerted evolution is thought to result from subsequent rounds of unequal crossing-over (or any other mechanisms fitting the "card deck" model (38)), which are favored by the tandemly repeated structure itself. The proposed unequal crossing-over mechanism, whether happening between sister chromatids or homologous chromosomes, explains the substantial degree of length polymorphism, here referred to as "expansion-contraction polymorphism," characterizing those sequences. Moreover, the ensuing shuffling of slightly divergent repeat units or Minisatellite Variant

Repeats (39) within the satellite generates additional internal site polymorphism. These peculiar properties of satellite sequences have made them an invaluable source of highly informative genetic markers, both in the human and in domestic species (reviewed in 31).

B. Detection of DNA Sequence Polymorphism

During the last ten years, a multitude of methods have been developed for the detection of DSP. Two techniques, however, undoubtedly dominate this field: Southern blot hybridization (40) and the Polymerase Chain Reaction or PCR (41), used either separately or in conjunction. A non-exhaustive list is reported here, the methods being grouped into four classes.

1. Restriction Pattern Analysis

DSP may alter the restriction patterns of defined chromosomal regions, generating so-called Restriction Fragment Length Polymorphisms (RFLP). Depending on the size-range of the explored restriction fragments, one will use either agarose gel electrophoresis, pulse-field gel electrophoresis (30) or polyacrylamide gel electrophoresis (42) for intermediate, large or small fragments respectively. RFLPs are classically detected by Southern blot hybridization. Alternatively, one can analyze restriction patterns of defined DNA sequences amplified by PCR, generating so-called Amplified Sequence Polymorphisms (43). When studying chromosomal rearrangements or expansion-contraction type polymorphisms, the use of PCR obviates the need for restriction enzyme digestion, the DSP reflecting itself in the size of the amplified product.

Because of its simplicity, the detection of RFLPs has by far been the most popular approach towards DSP. The relative lack of power inherent to the method (only 20% of a given sequence is amenable to exploration using the most common restriction enzymes) can be compensated

for by focusing on highly polymorphic sequences such as CpG dinucleotides (using enzymes such as TaqI and MspI containing CpG in their recognition sequence) or hyper-variable minisatellites. The discovery, however, of microsatellites as a very abundant source of highly informative DSP in a broad taxonomic range, easily detectable by PCR, is likely to shift the focus towards these sequences for future marker development (32-34, 37, 44).

2. Mismatch Analysis

Several methods for the detection of DSP are based on the study of mismatch analysis. DNA to analyze is probed with a sequence corresponding to a defined genetic variant. The presence of a different variant in the target DNA generates a mismatched heteroduplex, which can be detected by various means:

a. Detection of Altered Melting Behavior

A mismatched heteroduplex will differentiate itself from the perfectly matched homoduplex by an altered melting behavior which can be detected as an all-or-none, binary response: positive for the homoduplex, negative for the heteroduplex, or in a more graded response, allowing to distinguish between different heteroduplex variants.

The classical all-or-none test depends on the use of allele-specific oligonucleotides in hybridization experiments. By choosing appropriate hybridization and washing conditions, the allele-specific oligonucleotide will only recognize a perfectly complementary sequence (45). With the advent of PCR, new variants of this approach have been described including reverse dot-blot (46), the Amplification Refractory Mutation System (47) or allele-specific polymerase chain reaction (48), and Competitive Oligonucleotide Priming (49). The Ligation Amplification Reaction, amplification of specific DNA

sequences using sequential rounds of template-dependent ligation, can also be considered as a peculiar application of the allele-specific oligonucleotide approach (50).

More discriminating is Denaturing Gradient Gel Electrophoresis, exploring the pattern of melting behavior characterizing each heteroduplex when electrophoresed through an increasing gradient of DNA denaturants (51). The sensitivity of this method can also be improved by pre-amplifying the target sequence by PCR.

b. Ribonuclease and chemical mismatch detection

The presence of a mismatch in a heteroduplex makes those molecules susceptible to cleavage by various means including chemical treatment with either hydroxylamine or osmium tetroxide (52), as well as ribonucleases such as RNase A in case of an RNA:DNA heteroduplex (51). Electrophoretic analysis of the cleavage products allows one to distinguish different genetic variants. Again, implementing PCR will increase the sensitivity of the approach.

3. Single Stranded Conformation Polymorphism

Under nondenaturing conditions, single-stranded DNA has a folded conformation that is stabilized by intra-strand interactions. Consequently, the conformation, and therefore the electrophoretic mobility, is dependent on the sequence. DNA variants exhibit indeed mobility shifts when electrophoresed in such conditions, presumably resulting from conformational changes caused by sequence alterations, hence the name single stranded conformation polymorphism. Again, the altered mobility can be detected by blot hybridization analysis or relying on PCR (53, 54).

4. Direct Determination of the DNA Sequence

Obviously the most powerful approach towards DSP is the direct determination of the DNA sequence. The need of a cloning step, however, in classical sequencing protocols precluded the analysis of large samples. This limitation has been circumvented by the development of genomic sequencing (55), allowing the direct determination of defined DNA sequences from genomic DNA, and more recently and less laboriously by the development of direct sequence determination of PCR amplified products. The feasibility of the latter approach for the detection of DSP has been amply demonstrated in several independent studies (see, for instance, 56).

C. Origin and Evolution of DNA Sequence Polymorphism

DSP encountered in a given population find their origin in mutational events occurring in the germline and escaping the DNA repair machinery. The fate of these germline mutations in the population is dominated by two kinds of effects: stochastic and deterministic effects.

1. Stochastic Effects

When a new mutation appears in the population, its initial survival depends largely on chance, regardless of its selective effect. This is easily illustrated as follows. Assume an individual heterozygous for a neomutation inherited from its parent, in whose germline the mutation appeared. If this individual has one, two or three offspring, the chances for the neomutation to be lost from the population, because transmitted to none of the offspring, are 0.5, 0.25 and 0.125 respectively. Even if inherited by part of the offspring, the same "stochastic filter" will operate in the next generation. In the course of this random drift, the overwhelming majority of mutant alleles are lost by chance. However, some will see their frequency increase in the

population, and despite fluctuations over time, eventually become fixed in the population, until substituted by the next mutant allele.

As demonstrated essentially by Kimura (57) in the framework of his neutral theory of molecular evolution, the probability for a selectively neutral neomutation to be fixed in a population of N individuals is equal to its initial frequency 1/2N, the average time for fixation is four times the "effective" population size or 4Ne, and the rate k of mutant substitution per generation is simply equal to the rate of mutation per gamete and per generation, μ, independent of what the population size may be.

According to this view, a polymorphism observed in a population at a given time is composed of "transient" alleles catched in their stochastic "odyssey" throughout the population.

Populations for which $4.Ne.\mu \leq 1$ are essentially monomorphic, while populations for which $4.Ne.\mu \approx 1$ are characterized by a substantial degree of transient polymorphism. The model predicts a steady state level of heterozygosity, H:

$$H \approx \frac{4.Ne.\mu}{4.Ne.\mu + 1}$$

2. Deterministic Effects

There is evidence that the fate of a significant proportion of DSP, especially those occurring in non-coding parts of the genome (composing the large majority of the genome), is essentially dominated by random drift. However, when a neo-mutation affects a DNA sequence which is expressed at the phenotypic level in the broad sense, the mutation may not longer be selectively neutral, and deterministic effects will be superimposed on the stochastic ones. Negative and positive selection will respectively decrease or

increase the probability and rate of fixation, while "balancing selection" will maintain specific alleles in a population in an equilibrium state.

a. Negative Selection

When comparing DNA sequences between taxa, it appears that the estimated number of mutant substitutions per nucleotide to account for the observed divergence is highest for non-coding sequences, such as pseudogenes and intronic sequences, and much lower for coding sequences. For the latter, however, a difference must be made between first, second and third positions of the codons. The third position, for which only 28% of substitutions are expected to cause an amino acid change (versus 95% and 100% for first and second positions respectively), exhibits the highest substitution rate. When estimating that part of substitutions at the third positions which are so-called synonymous, rates very similar to non-coding regions are observed (57). Moreover, DSP are more prominent for non-coding sequences and, within coding sequences, at third codon positions (compared to first and second positions (58)). These observations are easily explained by assuming that the fate of neomutations arising in non-coding regions or of synonymous neomutations, is dominated by stochastic effects, while the fate of mutations causing amino acid replacements will depend as well on whether or not they disrupt the function of the protein, in which case they will be eliminated from the population by negative, "purifying" selection. The higher the functional constraints imposed on a protein, the higher the proportion of neomutations expected to be harmful and, hence, the lower the substitution rate, expressed at the protein level as a higher "unit evolutionary time" (average time required for one amino acid change to appear in a sequence of 100 amino acid residues).

These observations have been considered as a strong argument in favor of a predominant role for random drift in the dynamics of molecular evolution.

b. Positive selection

According to the previous discussion, the major drive behind molecular evolution is non-adaptive in nature, which is in conflict with the classical theory of adaptive, positive Darwinian selection.

There is, however, evidence for positive and adaptive evolution at the molecular level in at least a few instances. Comparing DNA sequences from members of two gene families: the serine protease inhibitors in rat (59) and the pregnancy specific ß1 glycoprotein gene family in man, evidence has been found for higher substitution rates at first and second codon positions than at the third position, in at least some protein regions, pointing towards positive selection.

Moreover, there are a number of experimental data suggesting that some allelic differences identified by electrophoresis are associated with adaptation to different environments. In Drosophila, for instance, their is evidence for correlation between in vitro heat resistance of ADH variants and the temperature characterizing their geographical origin (58).

c. Balancing selection

The evolutionary forces described so far generate transient DSP in the sense that the population frequencies of existing genetic variants will irrevocably change with time until either fixation or loss. In some cases, however, alleles may be maintained in a population at a steady state level. Overdominance is one of the mechanisms susceptible to generate such a "balanced polymorphism". For a two allele system, this means that the heterozygous individuals benefit from a selective advantage compared to both homozygous genotypes. This

is expected to generate a steady state where both alleles are maintained in the population at respective equilibrium frequencies p and q, where

$$q = \frac{s}{s + t} \quad \text{and} \quad p = \frac{t}{s + t},$$

s and t being the respective selection coefficients of the homozygotes.

The best known example of balanced polymorphism due to overdominance is the maintenance of the S α-globin allele (causing sickle-cell anemia in the homozygotes) as well as thalassemia-causing mutants (see, for instance, 61) in populations subjected to malaria, because of the resistance exhibited by the heterozygotes towards the parasite. The high level of polymorphism observed at the Major Histocompatibility Locus is thought to result from overdominance selection as well (62).

Frequency dependent selection may be another cause of balanced polymorphism, an example being the "rare mate advantage" observed in Drosophila (63).

II. CONSTRUCTION OF PRIMARY DNA MARKER MAPS

A. Linkage Strategies

Two loci are said to be genetically linked if, during meiosis, they recombine at significantly lower than a 50% rate, i.e., they generate significantly more parental gametes than recombinant gametes. The recombination rate between loci reflects the frequency of occurrence of an uneven number of crossing-overs between the loci. Because the probability for crossing-over is proportional to the distance separating the loci, the recombination rate can be used as a unit of chromosomal length. This length unit is known as the Morgan (M),

1 cM corresponding to the distance separating two loci exhibiting a 1% recombination rate. For small distance (≤30cM), the relation between centimorgan and recombination rate is essentially linear. For longer distances, however, the relation is more complex, depending on the frequency of double crossing-overs, itself affected by eventual interference.

Parental and recombinant gametes will only be distinguishable for doubly heterozygous individuals, hence the need for highly polymorphic markers.

Recently, and due to the advent of the PCR, it has been possible to directly determine the genotype of individual gametes (64). However, most of the time, the gametic contribution is inferred from the genotype of the offspring and linkage studies are performed within families. Most modern linkage studies use the lodscore test for evaluation of linkage: a sequential test based on the method of maximum likelihood (65). The lodscore corresponds to log10(LR), where LR corresponds to the ratio: likelihood of observations under alternative hypothesis $\theta \leq 0.5$, divided by the likelihood of observations under null hypothesis of no linkage, $\theta = 0.5$. In human genetics, a lodscore > 3 is accepted as significant evidence for linkage. The prior probability of linkage between two loci has been used to justify this stringent critical value. Note that 2ln(LR) can be used as well, having a chi-square distribution with one degree-of-freedom under the null hypothesis of no linkage.

Recently, algorithms for multilocus linkage analysis have been developed, allowing an estimate of the most likely gene orders and genetic distances between several loci simultaneously (66-68).

Although usually determined within families, genetic linkage can manifest itself at the population level also: a phenomenon called "linkage disequilibrium". According to the Hardy-Weinberg law, the

equilibrium genotypic frequencies are reached in a single generation (except if the initial gene frequencies are not equal among sexes). For a diallelic system with alleles a1 and a2, with respective allelic frequencies p1 and p2, the equilibrium genotypic frequencies are $p1^2$, 2p1p2 and $p2^2$ for a1a1, a1a2 and a2a2 respectively. This does not necessarily hold when considering two loci simultaneously. The genotypic equilibrium frequencies are only reached when the previous generation produces the four possible gametes at the expected frequencies: a1b1: p1q1, a2b1: p2q1, a1b1: p2q1, a2b2: p2q2. The difference between observed and expected gametic frequencies is called linkage disequilibrium, D. The value of D is reduced by d.θ every generation, θ being the recombination rate between the two loci. For unlinked loci D diminishes by 1/2 every generation; for linked loci, however, the reduction of D per generation will be much smaller. The detection of a linkage disequilibrium is an indication of linkage between the corresponding loci.

B. Genetic Maps

Using this linkage approach, combined with alternative mapping strategies such as "in situ" hybridization (see, for instance, 69), the use of somatic cell hybrid panels and radiation hybrid mapping (reviewed in 70) and comparative mapping (71), the map location of a large set of DSP can be determined in order to build a genetic marker map (see, for instance, 72-74). Assuming a total map length of 30M as for the human, and a desirable maximum distance of 20cM between markers, a set of 150 DSP could cover the entire genome. However, many more markers will be needed to generate reasonable maps for our domestic species, essentially for two reasons. First, most of the time we have no a priori information on the location of the characterized markers. Hence, some chromosomal regions will initially be

overrepresented in our map, others underrepresented. This problem is expected to become critical in the later stages of the development of a map. Comparative data will then become critical, allowing to search for markers whose location can be predicted from other species. Second, an individual will only be informative for the markers for which he is heterozygous; parts of his genome will thus not be explorable, because he will be homozygous for the corresponding markers. To compensate for this, one will have to identify more markers, the number required being inversely proportional to their heterozygosity -- hence, the importance of highly informative systems.

Once such a map is available, however, any gene for which the appropriate segregating family material is available can be located on the map. Assuming a maximum marker-target gene distance of 10cM, the expected lodscore for doubly informative, phase-known meioses approximates 0.16 (75). Therefore, 20 such meioses are theoretically sufficient to establish linkage with a lodscore of 3. In practice, however, the number of individuals to analyze will be higher, a function among other factors of the quality of the marker, expressed as its Polymorphism Information Content (76).

The efficiency of this approach has been illustrated by the recent mapping of a large number of genes involved in human single gene disorders (see, for instance, 77, 135). The identification of DNA markers for a defined gene can be the first step towards its molecular cloning. Successful "positional cloning", or the isolation of a gene based on its map location, has been achieved in the human for Chronic Granulomatous Disease, Duchennes Muscular Dystrophy, Retinoblastoma, Wilms Tumor, Cystic Fibrosis (134), Type-1 Neurofibromatosis and the Testis Determining Factor.

In domestic animals, genetic maps could be used to localize the genes underlying production traits,

allowing for Marker Assisted Selection, and a first step towards their isolation, the understanding of their mechanism of action and their manipulation by mutagenesis and gene transfer methods. Several laboratories around the world are now involved in the development of markers and the construction of genetic maps for our main domestic species, especially cattle, pigs and poultry.

III. GENETIC MAPPING OF QUANTITATIVE TRAIT LOCI

The majority of traits dealt with in animal production are so-called quantitative traits, characterized by continuous variation. The phenotype of an animal with respect to a particular trait is the result of the effect of a several "polygenes" known as Quantitative Trait Loci, or QTL, combined with environmental effects. The number of polygenes involved is essentially unknown. Classically, it is considered very large, each gene contributing a very small part of the genetic variation. However, there is evidence both from the plant world and the animal world, that QTL with significant effects are common (78, 79). The most likely model is to assume that there is indeed a large number of genes involved, but that there is a broad distribution of effects, substantial in some cases. Polygenes with extreme effects, whose segregation in a population may cause skewness and bi- or trimodality, are known as "major genes". Examples in animal breeding are "double muscling" genes in both cattle and pigs, the "White Shorthorn" gene involved in the determinism of "White Heifer Disease" and the "Booroola" fertility gene in sheep (80). Even with significant effects on the trait of interest, however, their contribution to the total genetic variation may be limited in case of low population frequency.

When dealing with quantitative traits, direct determination of genotype for the corresponding QTL is

impossible. Nevertheless, strategies have been designed to map QTL by linkage analysis. Within segregating populations, which is usually the case for our domestic species, QTL mapping can be performed both within families and at the population level.

A. QTL Mapping Within Families

Traditionally one proceeds as follows: offspring from an individual heterozygous for both marker and QTL are grouped according to which allele at the marker locus they inherited; a statistically significant difference between the phenotypic means of the two groups indicates linkage between marker and QTL. Test for statistical significance is done by linear regression (i.e. one-way analysis of variance) under the assumption of normally-distributed residual environmental variance. Classically, markers are tested one at a time for possible linkage with a QTL affecting the trait of interest. One of the drawbacks of this approach is that it is impossible to unequivocally estimate both map location of the QTL with respect of the marker, and its effect on the considered trait; no distinction can be made between a closely linked QTL with small effect and a loosely linked QTL with major effect.

Recently, the lodscore method has been improved, making it possible to deal with quantitative and other complex traits and fully exploiting the power of the nearly complete marker maps which have become available for different organisms. This approach is known as interval mapping. Not only does interval mapping solve the problem of simultaneous estimation of location and effect, but because of its increased power, it reduces the number of individuals to be tested to detect linkage with a QTL of given effect (81).

Assuming that the marker is the QTL, the number of individuals to test in order to detect an effect of given amplitude, δ, can be estimated from:

$$n \geq \frac{4(t_o+t_1)^2.s^2}{\delta^2}$$

where n gives the required sample size, s^2 is an estimate of the residual variance, t_o is the t value associated with Type I error, and t_1 is the t value associated with Type II error; t_1 equals tabulated t for probability 2(1-P) where P is the required probability of detecting δ if such a difference exists (82).

For dairy production for instance, and if performing the linkage analysis using the "daughter yield deviations" (DYD; $\sigma^2_{DYD} \approx$ 600 lb) from paternal half-sibs ("granddaughter design" (83), one would have to study 1,500, 378 and 168 individuals, respectively, to detect QTL with differences of 200 lb, 400 lb and 600 lb between alternate alleles. Assuming phenotypic variance of $(2500\ lb)^2$, such effects correspond to 0.08, 0.16 and 0.24 standard deviations, respectively. These estimates assume a Type I error of 5%, a Type II error of 10%, and absence of recombination between marker and QTL.

If the tested marker and the QTL recombine at a rate θ, the number of individuals to test increases by a factor $1/(1-2\theta)^2$ for single marker analysis, by a factor $\approx(1-r)/(1-2\theta)^2$ in the case of interval mapping, r corresponding to the recombination rate between the flanking markers (81).

In view of the costs and time involved in genotyping, it is important to minimize the required sample size. This can be achieved in various ways:

a. <u>Identification of the Individuals Most Likely to be Heterozygous, hence Informative, for the Studied QTLs</u>

One way to achieve this is to cross highly divergent strains for the trait of interest. In plant breeding, where the use of exotic germplasm is common practice, this is perfectly applicable. The

identification of markers for interesting QTLs from the exotic strains can then be used for their marker assisted introgression in the commercial varieties.

In animal breeding, however, introgression programs are very uncommon. With "Velogenetics" (described further below), however, the use of exotic germplasm in introgression programs may become more attractive for animal breeders as well.

An alternative approach is to identify the individuals whose offspring are showing a higher variance for the trait of interest.

b. Selective Genotyping of the Extreme Progeny

As pointed out by Lander and Botstein (75), the individuals whose genotype can be most clearly inferred from their phenotype are the ones providing most of the linkage information when studying complex traits. For quantitative traits, these are the individuals whose phenotypic value deviates most from the mean: the tails of the distribution. Sample sizes could be reduced by 60% and even 80% by focusing on individuals deviating one and two standard deviations, respectively, from the mean. Paradoxically, selective genotyping may be limited by the size of the studied population. Indeed, a larger sample will be required in order to find enough individuals one or two standard deviations from the mean.

c. Decreasing Environmental Variance via Progeny Testing

Weller et al. (83), have tested the effect of progeny testing to reduce the environmental variance by comparing the power of "daughter" and "granddaughter" designs for the detection of QTLs in dairy cattle. In the "daughter" design, marker genotype and quantitative trait values are assessed on daughters of heterozygous sires, while in the "granddaughter" design, marker genotypes are determined on sons of heterozygous sires,

their breeding values being determined by progeny testing from the quantitative trait value measured on their daughters. They demonstrate that for equal power the "granddaughter" design requires half as many marker assays as the "daughter" design.

d. Reducing genetic noise by searching for several unlinked QTL simultaneously, or "simultaneous" search (81).

e. Using DNA Pools

Instead of genotyping all individuals separately, one can analyze DNA pools from individuals sorted by phenotype. Significant differences of allelic frequencies between pools point towards possible genetic linkage between the corresponding marker locus and a gene or genes affecting the trait of interest. This approach can be used both for "within family" studies and for studies at the population level. The latter approach, however, requires linkage disequilibrium between QTL and marker locus. This method was first described by Arnheim et al. (84) to study the role of HLA class II loci in insulin-dependent diabetes mellitus. It was recently adopted by Plotsky et al. (85) to study association between DNA fingerprint bands and abdominal fat deposition in broilers.

f. Exploiting "Tagged QTLs"

The direct effect of selection for a production trait will be to increase the frequency of the favorable alleles at the segregating QTLs. However, this selection pressure may indirectly affect loci in linkage disequilibrium by so-called "hitch-hiking".

This is probably what happened to the genetic defect causing progressive degenerative myeloencephalopathy, or Weaver in Brown Swiss, shown to be linked to a major gene for milk production. Because of the

deleterious effect of the Weaver causing gene, it is the heterozygous "carrier" genotype which is selectively most advantageous, generating a "balanced polymorphism", the Weaver causing allele being maintained in the population at a relatively high frequency. This can be exploited, however, to map the corresponding QTL by going through the relatively easy exercise (compared to QTL mapping) of finding a marker linked to this single gene disorder. We have recently identified a marker linked to Weaver and presumably to the associated QTL.

Besides Weaver, QTLs for a variety of polygenic traits have been identified, both in plants and animals. Using complete DSP maps in tomato, Paterson et al. (78) identified at least six genes controlling fruit mass, four controlling soluble solids, and five controlling fruit pH, accounting for 58%, 44% and 48%, respectively, of the phenotypic variance. Martin et al. (79), using a similar approach, identified at least three tomato genes controlling water use efficiency. In cattle, Geldermann et al. (86) found significant effects on milk yield (+ 200 kgs) and fat content (+ 1%), especially for the ß-lactoglobulin locus. More recently, Cowan et al. (87) demonstrated significant effects on milk production traits using a prolactin DNA Sequence Polymorphism as marker.

B. QTL Mapping Within Populations

One can expect to find an effect of marker alleles linked to QTLs also outside of a family context, i.e., at the population level, if the two loci are in linkage disequilibrium. As reported by Hanset (88), and assuming a diallelic marker (alleles M1 and M2 with respective frequencies p1 and p2) linked to a diallelic QTL, the phenotypic difference between the respective homozygotes

at the marker loci is:

$$\delta = 2a \cdot \frac{D}{p1.p2}$$

with D measuring the linkage disequilibrium and 2a corresponding to the phenotypic difference between the two homozygotes for the QTL.

Markers for which a priori evidence for linkage disequilibrium is highest are the so-called "candidate genes": genes expected from their physiological role to be likely candidates for the QTL itself. DSPs at those loci, even selectively neutral by themself, can be expected to exhibit linkage disequilibrium with the hypothetical functional mutations because of their very tight linkage. As an example, the B allele of the K-casein gene has been shown in several studies to increase protein yield in milk by about 3%, and possibly to improve cheese yield independent of the effect on protein yield (see, for instance, 89, 90).

IV. USE OF DNA MARKERS IN BREEDING PROGRAMS

In classical selection programs, breeding values are estimated from individuals' own performances and performances of relatives (136). The expected genetic progress is a function of the accuracy of selection, i.e. the correlation between estimated and true breeding values. All direct information on QTL can be used to increase the accuracy of selection and, hence, genetic response. Early on, Soller and Beckmann (91) proposed to exploit marker information for the preselection of young dairy sires before progeny-test. In cattle, Marker Assisted Selection is already used for the sexing of preimplantation embryos using Y-specific probes (see, for instance, 92), and for genotyping at the K-casein (see, for instance, 93) and prolactin loci (87). In

pigs, Marker Assisted Selection is used to reduce the frequency of the major gene causing Porcine Stress Syndrome (PSS). Susceptibility to PSS correlates with Halothane sensitivity or Malignant Hyperthermia. This condition has been mapped to a linkage group on pig chromosome 6, encompassing the following markers: S(A-O)-GPI-Hal-H-A1BG-PGD (reviewed by 94). These markers are used for the Marker Assisted Selection against the PSS condition. Recently, the ryanodine receptor gene has been identified as a good candidate for the Malignant Hyperthermia or Hal gene (95).

As shown by Smith and Simpson (96), the gain to be made with Marker Assisted Selection increases with the proportion of QTL identified and is highest for low heritability traits. Unfortunately, the QTL determining the latter traits are also the hardest ones to identify. It should be noted that the increase in accuracy is subordinate to the accurate estimation of the QTL effects. This may require larger samples than the ones needed for the detection of linkage. Once a QTL mapped by within-family linkage studies, it may be more effective to identify supplementary flanking markers and to accurately determine the effect of the generated haplotypes at the population level. Selection can then focus on the best haplotype instead of spending initial selection efforts on intermediate ones.

The use of genetic markers in selection programs may as well reveal dominance deviation (particularly overdominance) and interaction deviation at defined QTL, variance components poorly dealt with in classical breeding theory. Specific programs may be required to fully exploit these QTL. In the case of overdominance for instance, two lines each homozygous for the different alleles at each QTL could be developed and crossed to produce multiple heterozygotes.

There is widespread interest in resolving quantitative traits into their Mendelian components by

mapping the underlying QTL. The implementation of marker assisted selection into breeding schemes, however, has not always been received with enthusiasm. Part of the skepticism expresses the doubt that the genetic gains obtainable by marker assisted selection will justify expensive and tedious large scale genotyping. Although the costs of genotyping will drop substantially in the near future, due to the rapid pace at which automation and robotics are being applied to DNA technology, this objection remains very pertinent.

Another major limitation of marker assisted selection under its present form, is its limitation to the exploitation of genetic variation preexisting within the commercial breed of interest, and only that present in a "high merit" genetic background. Favorable mutations appearing within a mediocre background, or present in "exotic" germplasm, would be difficult to exploit, even with markers.

We have therefore proposed a scheme, designed as "Velogenetics", combining marker assisted introgression and germ-line manipulations to reduce the generation interval, which might drastically increase the power of marker assisted selection (141).

IV. INDIVIDUAL IDENTIFICATION AND PATERNITY DIAGNOSIS:

Methods to estimate the breeding value of an animal use information from relatives. As a matter of fact, keeping track of familial relationships has always been one of the major concerns of animal breeders, and parentage control is now a widely used procedure for several domestic species. Parentage control relies on the use of polymorphic systems within the studied population. The alleles that characterize an individual originate from the mother or the father. If one of the parents is known (usually the mother), the alleles necessarily transmitted by the other parent can be

deduced easily. Paternity testing consists of scoring the existence or lack of those obligate paternal alleles in the genotype of the putative parent. Lack of one or more of these alleles points towards incorrectly assigned paternity. If, on the contrary, all obligate paternal alleles are present in the tested parent, there is no evidence for incorrectly assigned paternity. Nevertheless, one always has to consider the possibility of fortuitous coincidence. The higher the variation of the genetic markers used, the higher the probability to detect incorrectly assigned paternity, thus the higher the "exclusion power".

Until now, the systems most often used for paternity testing were blood group systems, biochemical polymorphisms, or the major histocompatibility system. The availability of DSP, however, opens new perspectives for paternity diagnosis. Hypervariable minisatellites in particular, characterized by their remarkably high degree of polymorphism, have proven especially useful in this respect. Multilocus DNA fingerprints, based on the simultaneous detection of related minisatellite loci, have been shown to be extremely powerful for paternity diagnosis, both in human (19) and animals (108, 109). Exclusion powers as high as 99.999996% have been obtained with as few as 2 probes in the human (19). With such high exclusion powers, absence of exclusion can be considered proof for true biological parentage. Another corollary is that very high exclusion powers can be obtained even when a single parent is available and tested for parenthood. Multilocus DNA fingerprints, however, tend to be replaced by the combined use of a limited number of locus-specific VNTR markers (20), giving equally powerful, but more reproducible, sensitive and easily interpretable patterns. With the advent of locus-specific VNTRs and PCR-amplifiable microsatellites in animal species (44), the same will probably hold in this field too.

Along the same lines, DNA markers can be used as well for individual identification. Using expansion-contraction type polymorphisms, individual specific "DNA bar codes" can literally be generated (19, 110).

SUMMARY OF THE INVENTION

Disclosed herein is a set of locus-specific genetic markers for domestic cattle and related bovids, that constitute a primary bovine DNA marker map. Among other applications, these markers and the map are useful for:

- individual identification,
- parentage testing,
- the genetic mapping of economic trait loci, or genes involved in the determinism of economicalli important traits, whether single gene traits or complex multifactorial traits,
- marker assisted selection,
- velogenetics, or the synergistic use of marker assisted introgression and germ-line manipulations to reduce the generation interval.

The usefulness of this set of markers for the genetic mapping of economic trait loci is illustrated by the identification of a genetic marker for bovine progressive degenerative myelo-encephalopathy or "Weaver" in the Brown Swiss breed.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1: shows a typical VNTR pattern obtained with probe GMBT-005, using HaeIII.

Figure 2: Example of a microsatellite pattern (TGLA9).

Figure 3: Schematic representation of "Velogenetics".

DETAILED DESCRIPTION OF THE INVENTION

I. CONSTRUCTION OF A PRIMARY BOVINE DNA MARKER MAP:

Our laboratory has focused in the last two years in the development of a primary DNA marker map for cattle. We have now developed more than 300 highly polymorphic DNA markers of either of three types:

1. Variable Number of Tandem Repeat Markers (VNTR)

Hypervariable minisatellites are known to show significant cross-hybridization between species (31, 44, 110). We have exploited this to isolate bovine VNTRs using heterologous minisatellite probes. Screening purpose-built libraries with minisatellite probes, we have isolated 36 bovine VNTRs, characterized by a mean heterozygosity of 59.3% within the American Holstein breed. Matching probabilities and exclusion powers were estimated by Monte-Carlo simulation, showing that the top 5 to 10 probes could be used as a very efficient DNA-based system for individual identification and paternity diagnosis. The isolated VNTR systems should contribute significantly to the establishment of a bovine primary DNA marker map. Linkage analysis, use of somatic cell hybrids and in situ hybridization demonstrate that these bovine VNTRs are organized as clusters, scattered throughout the bovine genome, without evidence for proterminal confinement as in the human (35). Moreover, Southern blot analysis and in situ hybridization demonstrate conservation of sequence and map location respectively of minisatellites within Bovidae. A typical VNTR pattern obtained with one of our probes is shown in Figure 1. Detailed description of our VNTR systems is reported in "Example 1".

2. Multisite haplotypes

We used 110 random cosmids to probe Southern blots of 9 unrelated cattle DNAs digested with 12 restriction enzymes. Although only one third of the expected fragments could be detected, 85% of the cosmids revealed at least one polymorphism. The mean heterozygosity of the generated multisite haplotypes (98) was estimated at 51.9% . A surprisingly high proportion of polymorphisms (≈25%) was attributed to insertion-deletion events, compensating for the lower level of nucleotide diversity, π, observed in cattle ($\pi \approx 0.0007$) as compared to the human. The mutation rate at cytosines in the CpG dinucleotide was estimated approximately 10 times higher compared to other nucleotides. The generated markers should cover approximately 40% of the bovine genome when used in linkage studies. A detailed description of our multisite haplotypes is reported in "Example 2".

3. Microsatellites

Recently, microsatellites were proven to be an abundant source of highly polymorphic markers in the human (32-34). As their name implies, microsatellites are minute VNTR markers (18-20), characterized by tandem repetitions of very short repeats, one to four base pairs in length. Microsatellites exhibit levels of polymorphism comparable to VNTRs, but are much more abundant and apparently evenly spread throughout the genome. We have estimated the frequency of (CA)-dinucleotide repeats in the bovine genome at ≥ 150,000. Because of their small size, their detection is greatly facilitated by PCR. Although this imposes the preliminary determination of flanking DNA sequences to design the appropriate primers, the subsequent PCR reaction used for their analysis offers several advantages over Southern blotting, being fast, requiring less DNA and being easier to automate.

As part of our effort to build a primary DNA marker map for cattle, we have isolated more than 250 bovine microsatellites, amplified most of them in vitro and shown that the majority of them are indeed polymorphic in cattle. Several of these have been tentatively assigned to specific bovine chromosomes using a somatic cell hybrid panel. Moreover, we have shown that approximately 50% of the bovine microsatellites can be successfully used in other Bovidae as well, which will greatly facilitate the construction of marker maps in these species.

Magnetic solid phase DNA sequencing procedures (137) are used for the massive generation of sequence information and multiplex approaches for genotype collection, based on the simultaneous detection of molecules labelled with different fluorescent dyes using a laser-excited confocal fluorescence gel scanner (139).

A typical microsatellite pattern is shown in Figure 2. A detailed escription of our microsatellites is reported in "Example 3".

The relative location of the markers was determined by linkage analysis in pedigrees generated by multiple ovulation and embryo transfer. To assign linkage groups to specific chromosomes, highly polymorphic "anchor markers" were mapped using somatic cell hybrids (Jim Womack, Texas A&M), and by in situ hybridization (Rudy Fries, ETH - Zurich).

Linkage analysis involving 150 of these markers, generated a primary DNA marker map with 24 linkage groups counting two or more markers (15 assigned to specific chromosomes or synteny groups), and 68 singleton markers. A detailed description of our primary bovine DNA marker map is reported in "Example 4".

II. MICROSATELLITE MAPPING OF A MAJOR GENE FOR MILK PRODUCTION, LINKED TO BOVINE PROGRESSIVE DEGENERATIVE MYELOENCEPHALOPATHY OR WEAVER.

Identifying polygenes, requires the analysis of pedigrees of considerable size, despite the development of procedures such as interval mapping, simultaneous search, selective genotyping, etc. In this work we have explored an alternative approach to map a polygene, exploiting the association observed in cattle between the single gene disorder "Weaver", and increased milk production. Weaver or bovine progressive degenerative myeloencephalopathy is a recessive disorder characterized by the appearance between 5 and 8 months of age of bilateral hind leg weakness, ataxia with deficient proprioceptive reflexes, without skeletal or muscular defects. Estimates of gene frequency in the American Brown Swiss breed point towards the maintenance of the Weaver gene at relatively high frequency (≥5%), despite the implementation of programs for detection and elimination of carrier bulls. Moreover, Hoeschele and Meinert (140) showed that Weaver carrier animals have an advantage of 690.8 kgs milk (> 0.25 phenotypic σ) above the mean. Both observations could be accounted for by the presence of a gene with major effect on milk yield in linkage disequilibrium with the "weaver" gene.

Brown Swiss animals showing symptoms of Weaver were identified with the help of the American Brown Swiss Association. Blood samples were collected from the affected animals, their parents, and full-siblings when available. Diagnosis of Weaver was confirmed in most cases by anatomopathological examination of spinal cord and cerebellum at the Department of Pathology of the College of Veterinary Medicine, Kansas State University. Shrunken Purkinje cells in the cerebellum, spheroids and degenerated myelin sheets in the spinal cord were considered pathognomonic. Altogether, 78 animals were

identified generating a single, large pedigree. All animals were genotyped for more than 70 genetic markers: 40 Variable Number of Tandem Repeat markers and more than 30 Microsatellites. Linkage analysis was performed using the "LINKAGE" programs (60). The microsatellite marker TGLA116 was giving a highly significant lodscore of 6.5 for a recombination rate of 7.5%. Although a priori probability for pair-wise linkage is unknown in cattle, a lodscore of 3 is generally considered to be the threshold for statistical significance as in the human. This value (5.8) was obtained assuming complete penetrance. Actual penetrance for the Weaver condition is unknown. However, and because our pedigree was constructed by sampling clearly affected animals, the assumption of complete penetrance is very reasonable in this situation.

The marker TGLA116 is characterized by three alleles segregating in our Weaver pedigree. 72% of the affected individuals were of the 3/3 genotype, 16% of the 2/3 genotype, and 12% of the 1/3 genotype. Hence, and at least in our family material, the "Weaver" allele was clearly associated with allele 3 at the marker locus. Whether similar disequilibrium will be observed at the population level remains to be determined. The reported lodscore values were obtained using allelic frequencies estimated on a sampl of 135 sires from the American Brown Swiss breed.

Because of the biased sampling procedure used to generate the pedigree markers showing distorted segregation could generate erroneous evidence for linkage with the disease. A "control" pedigree, consisting of more than 100 Weaver-free Holstein individuals, was therefore typed for TGLA116 as well. The microsatellite marker was characterized in this pedigree by the same three alleles, with respective frequencies of 18%, 57% and 25% for alleles 1, 2 and 3, showing a perfect Mendelian segregation. Therefore, it is concluded that marker

TGLA116 is genetically linked to Weaver. From the generated lodscore curves, the genetic distance between the two loci is estimated at 3 ± 10 centiMorgan. The limits of this 95% confidence interval correspond to recombination rates with lodscores one unit below the obtained maximum lodscore. Because of the tight linkage between TGLA116 and Weaver, this marker should be linked to the associated QTL as well. The distance between TGLA116/Weaver and QTL is, however, unknown at this point. The effect using Weaver as marker, however, was of such magnitude that the genetic distance separating these loci is unlikely to be great. We are in the process of determining the relative location of these three loci.

In consequence, the TGLA116 marker will allow us to perform marker assisted selection against the Weaver condition. Indeed, it is now possible for offspring from individuals heterozygous for both the Weaver condition and TGLA116, to estimate the genotypic likelihoods at the Weaver locus based on their TGLA116 genotype and that of their parents.

In addition, we are now in a position to test the effect of the corresponding chromosomal segment on milk production.

III. VELOGENETICS

Few question the fundamental interest of resolving quantitative traits into their Mendelian components by mapping the underlying QTL. The implementation of Marker Assisted Selection into breeding schemes, however, has not always been received with a lot of enthusiasm. Part of this skepticism reflects the disbelieve that DNA Marker Maps will become available for our domestic species within a reasonable time-span, or that QTL can be identified by linkage strategies. In our view, these arguments only reflect the lack of information of their

protractors. On the other hand, part of the skepticism expresses the doubt that the genetic gains obtainable by Marker Assisted Selection will justify expensive and tedious large scale genotyping. Although the costs of genotyping will drop substantially in the near future, due to the amazing pace at which automation and robotics are applied to DNA technology, this objection remains quite pertinent.

Another major limitation of Marker Assisted Selection under its present form, is its limitation to the exploitation of genetic variation preexisting within the commercial breed of interest, and only if present in a "high merit" genetic background. Favorable mutations appearing within a mediocre background, or present in "exotic" germplasm, would be difficult to exploit, even with markers.

We propose a scheme, combining Marker Assisted Introgression and germ-line manipulations, to reduce the generation interval -- which might drastically increase the power of Marker Assisted Selection: "Velogenetics".

A. Marker Assisted Introgression

The basic principle underlying Marker Assisted Introgression are well-known. A gene responsible for a favorable attribute can be introgressed from a "donor" strain into a "recipient" strain by repeated backcrossing. During the introgression process, the retention of the favorable gene is monitored in the backcross products, with linked, flanking DNA markers. This latter aspect is particularly important for traits involving multiple genes and/or characterized by sex- or age-limited expression. Classical genetic theory tells us that, with the exception of the "marked" segment whose retention is desired, the genomic contribution of the donor line is diluted by half after each backcross. Hence, and after four backcrosses, the recipient genome is reconstituted to ± 90% of the original. At the

marked locus, however, the backcross retains one copy of the desired "donor" variant. If required, one intercross will then generate 25% of offspring homozygous for the favorable donor variant. The net result is a "graft" of an advantageous gene within a recipient background. The procedures entirely respects organization and chromosomal localization of the grafted gene, avoiding aberrant expression patterns, which are too often characterizing transgenes.

Noteworthy, the gene to be transferred does not need to be cloned per se. Only its genetic map location is required, as defined by the availability of linked markers, ideally flanking the gene of interest on each side. Hence, this procedure is perfectly applicable for the introgression of QTL identified through the previously described mapping strategies.

Marker Assisted Introgression can easily be applied to several genes simultaneously. This feature will be of particular interest for complex traits involving several genes. Introgressing more than one gene from a donor to a recipient line, however, increases the selection intensity at each backcross: with 1 marker, 1/2 of the offspring have the favorable genotype, with 2, 1/4 and with n markers, $(1/2)^n$.

Selecting for the retention of defined "donor" genes will hamper the recovery of the recipient background genotype in adjacent chromosomal regions. This can be compensated for by increasing the number of backcrosses, or better by monitoring the fate of additional adjacent markers to identify the backcross products resulting from recombinations as close to the "grafted" gene as possible.

B. Shortening the Generation Interval of Domestic Species by "Velogenesis"

Introgression by repeated backcrossing, assisted or not by genetic markers, is common practice in a variety

of organisms, but is essentially unfeasible in domestic animals such as cattle, because of their prohibitively long generation time. The generation interval of such species could, however, be reduced based on the "in vitro" maturation and fertilization of foetal oocytes, hereinafter referred to as "Velogenesis".

An overview of female gametogenesis (100,101), indicates that the feasibility of such a scheme may not be that far-fetched. Briefly, oogenesis begins with the formation of primordial germ cells in the region of the allantois. These precursor cells migrate to the developing gonads where after a period of mitotic proliferation, they enter meiosis. Meiosis is arrested at the diplotene stage of prophase I by the poorly understood "meiotic division I arrest system", after which the primary oocyte enters a resting phase. During the life time of the animal, small numbers of resting primary oocytes are successively recruited into a pool of growing oocytes, within the environment of a gonadotropin-dependent developing follicle. These activated oocytes growth in size, acquire the competence to resume meiosis if appropriately stimulated, and accumulate the required material to sustain the early stages of the subsequent embryonal development. Resumption of meiosis and oocyte maturation is triggered by a hypothetical maturation-inducing signal produced by granulosa cells in response to gonadotropins. At least in rodents, oocyte maturation seems to be mediated by a drop in cyclic AMP in the oocyte and subsequent inactivation of a type A protein kinase. Evidence for the role of this pathway in oocyte maturation is, however, much more controversial in ruminants. Note that in the granulosa cells, gonadotropins act, among other pathways, through the activation of adenylate cyclase with subsequent increase in cAMP concentrations (102). In the oocyte, a cascade of still to be determined events then probably leads to the phosphorylation

and activation of a phosphatase, probably homologous to the S. Pombe cdc25 gene (103), which will itself dephosphorylate and activate the M-phase promoting factor (MPF), now known to be a complex of a $p34^{cdc2}$protein kinase subunit with a B type cyclin (see 104 for a review). The maturing oocyte completes the first meiotic division and enters the second (becoming a secondary oocyte) which will be arrested as well at metaphase II until fertilization. This "meiotic division II arrest system" is thought to reflect the stabilization of MPF mediated by the kinase activity of $pp39^{mos}$ on either a cyclin protease or on cyclin itself. Fertilization relieves this block, by increasing the intracellular Ca^{2+} concentration, triggering calcium-dependent protease activity (reviewed in 104).

In cattle, primordial germ cells reach the genital ridge at about 40 days of gestation. After a period of mitotic proliferation, they differentiate into oogonia starting around 60 days of gestation. Mitotic proliferation of the germ line ceases around day 170 of gestation fixing the maximum number of oocytes the female will ever have. Meiosis starts at about 80 days, and the first primordial follicles are discernable at 90 days of gestation. Remarkably, activation of resting primordial follicles starts already <u>in utero</u>, around day 140, and secondary and tertiary follicles can be seen at 210 and 230 days, respectively. It is estimated that 2 to 4 resting primordial follicles are recruited daily into the pool of activated, developing follicles. These activated foetal oocytes, however, are irrevocably committed to follicular atresia. Indeed, spontaneous oocyte maturation and ovulation do not begin until puberty. Submitted to appropriate hormonal stimulus, however, prepubertal oocytes can resume meiosis, can be fertilized and can produce viable offspring. Indeed, offspring have been obtained from gonadotropin-stimulated calf oocytes, transferred to postpubertal

recipient animals (reviewed in 105). The purpose of velogenesis would be to attempt to obtain similar results with foetal oocytes at the earliest stage possible, as early as 90 to 180 days of gestation.

Very encouraging is the development in mice of culture systems supporting the growth of primary follicles, yielding mature oocytes capable of fertilization in vitro and development to term (106, 107). It is reasonable to anticipate that similar conditions, supporting development of bovine oocytes, will become available in a species were primary oocytes from relatively small antral follicles can already be successfully matured and fertilized in vitro.

On way to achieve velogenesis would be to attempt to rescue oocyte nuclei from primordial follicles by their transfer into enucleated, maturable oocytes.

So far we have only discussed velogenesis through the reduction of the female generation interval. "Male" velogenesis could similarly be accomplished by the early stimulation of spermatogenesis.

The impact on breeding programs of "velogenesis" or the reduction of generation time by in vitro maturation and fertilization of fetal oocytes has been discussed by Betteridge et al. (101). In dairy breeding, for instance, annual responses in milk yield could be doubled compared to conventional progeny testing. With the added power of Marker Assisted Introgression, the approach becomes much more powerful. "Velogenetics", or the synergistic use of Marker Assisted Introgression and "velogenesis", can be viewed as a procedure for the rapid and efficient intraspecies transfer of desirable genes between genetic backgrounds. By analogy with the term "transgene", the manipulated genes are referred to as "velogenes".

In particular, desirable traits identified outside commercial breeding stock, could be quickly introgressed into high merit genetic backgrounds. Examples would

include disease resistance, genes affecting milk and meat composition, Polled, coat color genes, etc. Moreover, the possibility to exploit "exotic" genetic variation identified outside the breed of interest is particularly attractive because it greatly facilitates the mapping of the genes of interest.

A schematic representation of "Velogenetics" is shown in Figure 3.

The present invention is further detailed in the following Examples, which are offered by way of illustration and are not intended to limit the invention in any manner. Standard techniques well known in the art or the techniques specifically described below were utilized.

EXAMPLE 1

CHARACTERIZATION OF A SET OF VARIABLE NUMBER OF TANDEM REPEAT MARKERS CONSERVED IN BOVIDAE.

INTRODUCTION

Human minisatellite sequences, exhibiting very high levels of genetic polymorphism due to variation in the number of tandem repetitions, have proven an invaluable source of genetic markers commonly termed "VNTRs" (18-20). VNTRs have been instrumental in the genetic mapping of several disease-causing genes, as tools for individual identification and paternity diagnosis and to address a variety of biological issues, including imprinting, loss of heterozygosity in malignancies, etc.

In animal genetics, highly polymorphic markers such as VNTRs could similarly be used for individual identification and paternity diagnosis - relying today on less informative biochemical polymorphisms and blood

group systems -, and for the mapping of so-called economic trait loci (ETL) or genes involved in the determination of production traits. Classically, artificial selection has relied on the biometrical evaluation of breeding values from individual performance records and from performances of relatives (136). One of the powers of the biometrical approach is that it obviates the need for any detailed molecular knowledge of the underlying genes or ETL. However, it is believed that the genetic mapping of ETL could be used to increase genetic response by affecting accuracy and speed of selection, through a procedure called marker assisted selection (MAS) (91, 96). Moreover, defined alleles could be moved efficiently between genetic backgrounds by velogenetics or the combined use of marker assisted introgression and germline manipulations aimed to reduce the generation interval (141, 142).

We report the cloning and characterization of 36 bovine variable number of tandem repeat (VNTR) markers, characterized by a high degree of polymorphism within commercial herds and shown to be conserved within Bovidae.

MATERIALS AND METHODS

1. Cloning of bovine VNTRs and detection of polymorphism:

500μg genomic DNA from 20 unrelated cows was digested to completion with MboI or HaeIII. After two rounds of size fractionation by agarose gel electrophoresis, electroelution and addition of EcoRI linkers (only for HaeIII restricted DNA), fractions from 3 to 4 Kb (kilobases), from 4 to 6 Kb and above 6 Kb were ligated into the BAP-dephosphorylated BamHI and EcoRI sites, respectively of pUC13. Approximately 80,000 independent clones were obtained by

transformation of DH5α cells, and were screened successively with the following minisatellite sequences: the 282 base pair HaeIII-ClaI fragment containing the minisatellite in the protein III gene of wild-type M13, pUCJ, pSP64.2.5EI ('Per'), pα3'HVR64, pINS310, EFD134.7 and pS3 (20, 21, 110, 123, 143). Hybridization and washings were done in the conditions used to generate multilocus DNA fingerprints with the respective probes (110, 143). To check for polymorphism, plasmid DNA isolated from positive colonies was used to probe MboI, HaeIII and TaqI Southern blots of 18 randomly selected American Holsteins. Hybridizations were done at 65 C in 7% SDS, 10% PEG, 50mM NaHPO4 with addition of 50μg/ml bovine genomic DNA. Final washes were at 65 C in 0.1xSSC, 0.1% SDS. When using bovine probes on ovine Southern blots, hybridization and washing temperatures were reduced by 10 C.

2. Estimation of Matching Probabilities and Exclusion Powers:

Allelic frequencies were estimated from the sample of 18 randomly selected American Holsteins. Matching probabilities and exclusion powers (113) were then estimated by Monte-Carlo simulation (10,000 simulations in each case), assuming Hardy-Weinberg equilibrium and using "Pat-Power", a program designed by one of us. The following parameters were estimated: MPR: matching probability for two randomly selected individuals; MPS: matching probability for full-sibs; EPR: probability to exclude an alleged father unrelated to the real one (mothers phenotype known); EPS: probability to exclude an alleged father full-sib to the real one (mothers phenotype known); EP1: probability to exclude a wrongly assigned parent without phenotypic information available from the other one.

Patpower calculates Matching Probabilities and Exclusion Powers characterizing given autosomal polymorphic systems, by Monte-Carlo simulation.

Matching Probabilities relate to individual identification and express the likelihood that two individuals would have the same pattern with a given probe. Patpower calculates two types of Matching Probabilities: MPR, the Matching Probability for two unrelated individuals, and MPS, the Matching Probability for two full-sibs.

Exclusion Powers relate to paternity diagnosis and express the likelihood that a wrongly assigned paternity or maternity will be detected with a given probe. Patpower determines three types of Exclusion Power: EPR, where one parent is known with certainty, the proband is unrelated to the other real parent; EPS, where one parent is known with certainty, the proband is full-sib of the other real parent; and EP1, whre only the proband is available.

The user needs to input the number of alleles characterizing the polymorphic system in the population of interest, their respective frequencies, and their dominance-recessivity relationships. For the ABO blood group system in humans, for instance, A and B are codominant and both dominate O. Each allele is given a binary code following the rules of the "LINKAGE" program (60).

Patpower then stochastically generates a pair of parents with an offspring, a full-sib of the real father and an unrelated individual. "Phenotypes" are obtained from the genotype using the boolean "or" operator and are used to determine matching between unrelated individuals (MPR) and between full-sibs (MPS), as well as the exclusion of the unrelated individual considered as a proband, with (EPR) and without information (EP1) from one of the real parents, and exclusion of the uncle considering information from the real mother (EPS). This simulation is repeated as many times as determined

by the user, allowing for the estimation of the respective livelihoods.

3. Segregation and linkage analysis:

All members of an American Holstein pedigree with 91 offspring obtained by multiple ovulation and embryo transfer (MOET) from 20 parents, were genotyped for all identified markers. Each parent has a mean of 9.1 offspring with a mean of 1.9 partners. Segregation and linkage analysis were done with slightly modified versions of the "LINKAGE" programs as previously described (31).

4. Synteny mapping:

The hybrid somatic cells were prepared by fusion as previously described (97). Southern blot hybridization and concordancy analysis were done according to Threadgill et al. (114).

5. In situ hybridization:

Chromosomes were prepared as described by Fries et al. (115) and chromosome identification was based on QFQ-banding and according to the international standard (116). Probe preparation and in situ hybridization were as previously described (144).

RESULTS

1. A set of bovine VNTR markers:

Using the strategy described above, we have isolated a total of 36 bovine VNTRs, listed in Table 1. Polymorphic patterns were attributed to minisatellite sequences when characterized by more than two alleles distinguishable with more than one restriction enzyme. Seven additional, non VNTR-type polymorphisms were detected during this experiment and are reported as well.

TABLE 1

VNTR Clones

Name	Locus[1]	Polymorph[2]	Enz.[3]	Het.[4]	Ovin[5]
GMBT-002	DY1651b	VNTR	HaeIII	52	P
GMBT-003		VNTR	HaeIII	56	
GMBT-005	D24Sib	VNTR	HaeIII	85	N
GMBT-006	D14Sib	VNTR	HaeIII	73	N
GMBT-007	DU10Sib	VNTR	HaeIII	96	P
GNBT-008		VNTR	TaqI		P
GMBT-009	DU22Sib	VNTR	HaeIII	58	
GMBT-011	D26Sib	VNTR	HaeIII	85	P
GMBT-012		VNTR	MboI	22	
GMBT-013		VNTR	HaeIII	4	
GMBT-015	D21S3b	VNTR	HaeIII	61	P
GMBT-016	D21S12b	VNTR	HaeIII	78	P
GMBT-017	D8Sib	VNTR	HaeIII	15	M
GMBT-019	D10Sib	VNTR	MboI	7	
GMBT-020		VNTR	MboI	65	
GMBT-021	D21S2b	VNTR	HaeIII	65	
GMBT-022	D18Sib	VNTR	MboI	40	
GMBT-025		VNTR	HaeIII	25	
GMBT-026		VNTR	HaeIII	26	
GMBT-027		VNTR	MboI	40	
GMBT-028	D2Sib	VNTR	HaeIII	81	
GMBT-031		VNTR	HaeIII	58	
GMBT-033		VNTR	HaeIII	70	
GMBT-034		VNTR	HaeIII	20	
GMBT-035		VNTR	HaeIII	59	
GMBT-036	DU27Sib	VNTR	HaeIII	89	
GMBT-039		VNTR	HaeIII	33	
GMBT-041	D23Sib	VNTR	HaeIII	81	
GMBT-042		VNTR	HaeIII	78	
GMBT-047	D2S2b	VNTR	HaeIII	65	
GMBT-049		VNTR+PM	HaeIII, Mbo		

TABLE 1 (Continued)

VNTR Clones

Name	Locus[1]	Polymorph[2]	Enz.[3]	Het.[4]	Ovin[5]
GMBT-051		VNTR	HaeIII	94	
GMBT-053		VNTR	HaeIII	59	
GMBT-058		VNTR	TaqI	89	
GMBT-059	DU10S2b	VNTR	BamHI	67	
GMBT-060		VNTR+PM	MspI	87	
GMBT-002		PM	TaqI	37	
GMBT-024		PM	TaqI	62	
GMBT-029		PM	MboI	28	
GMBT-014	DU22S2b	(?)	MboI, TaqI	68	
GMBT-018		(?)	TaqI	17	

---

1 LOCUS:locus name following HGM nomenclature rules whenever available from mapping studies.

2 POLYMORPH:type of polymorphism (VNTR:Variable Number of Tandem Repeats; PM:Point Mutation; (?):unexplained).

3 ENZ:preferred restriction enzyme for its detection.

4 HET:heterozygosity within Holszteins, estimated from a sampel of 27 presumably unrelated Holstein animals.

5 OVIN:cross-reaction in sheep; N, negatie; M, monomorphic; P, polymorphic, not tested.

Within the American Holstein breed, the mean heterozygosity over all VNTR systems was 59.3%. When using probe GMBT-016 with MboI instead of HaeIII, and supposedly because of the presence of minisatellite variant repeats (MVR) (39) harboring MboI sites, an extremely variable, locus-specific "midisatellite" pattern (36) is generated (data not shown). Used with MboI, this probe is particularly powerful for individual identification.

We found one clear instance of maternal neomutation with probe GMBT-022. Besides this, all probes showed proper Mendelian segregation.

Table 2 reports estimated matching probabilities and exclusion powers as well. Systems GMBT-009, GMBT-011 and GMBT-022 were treated as "open" systems, meaning that - because of their small size - some alleles were not detectable in our conditions. To avoid ambiguities in identification and paternity diagnosis, these unidentified alleles were pooled in a single "recessive" class. For individuals showing a single band, no distinction was made between homozygosity and heterozygosity based on band intensity.

Discrepancies between probe ranking according to heterozygosity versus ranking according to matching probabilities and exclusion power, most probably results from the small sample size used to estimate both types of parameters. However, heterozygote advantage at some loci could be an alternative although unlikely explanation in view of the apparent neutral behaviour of human minisatellite sequences (124).

TABLE 2

Matching Probabilities for VNTR Clones

Name	MPR[1]	MPS[2]	EPR[3]	EPS[4]	EPI[5]
GMBT-002	21.24	52.05	36.27	17.31	19.66
GMBT-003	24.25	53.53	30.90	15.37	16.56
GMBT-005	12.02	43.24	49.53	24.18	31.99
GMBT-006	19.38	49.49	36.93	18.28	20.56
GMBT-007	03.25	33.38	72.45	34.97	56.19
GNBT-008	--	--	--	--	--
GMBT-009	12.13	43.73	45.86	21.84	30.42
GMBT-011	10.88	42.14	49.34	24.34	33.24
GMBT-012	44.09	68.44	17.95	08.87	06.45
GMBT-013	92.28	96.28	01.96	00.84	00.16
GMBT-015	12.99	45.00	47.48	22.82	29.82
GMBT-016	12.59	44.10	48.87	24.15	30.17
GMBT-017	64.57	81.50	09.92	04.84	22.60
GMBT-019	42.29	96.26	01.91	00.89	00.07
GMBT-020	48.29	70.38	13.83	07.61	05.78
GMBT-021	21.22	50.93	35.26	17.79	19.21
GMBT-022	02.39	32.78	75.43	37.23	61.60
GMBT-025	50.37	72.51	14.04	07.47	05.04
GMBT-026	58.10	77.26	13.16	06.39	03.11
GMBT-027	--	--	--	--	--
GMBT-028	03.19	32.60	74.11	36.45	58.33
GMBT-031	20.82	50.89	36.05	17.81	20.19
GMBT-033	--	--	--	--	--
GMBT-034	71.41	84.57	07.72	03.74	01.44
GMBT-035	24.24	53.17	32.31	15.86	16.80
GMBT-036	04.19	35.42	69.74	33.07	53.05
GMBT-039	39.79	63.90	17.83	08.65	10.16
GMBT-041	12.18	43.84	48.09	22.72	30.41
GMBT-042	07.38	38.85	57.61	27.43	40.02
GMBT-047	30.09	55.38	25.43	12.56	14.79
GMBT-049	--	--	--	--	--

TABLE 2 (Continued)

Matching Probabilities for VNTR Clones

Name	MPR[1]	MPS[2]	EPR[3]	EPS[4]	EPI[5]
GMBT-051	02.40	32.23	76.37	38.01	61.37
GMBT-053	15.16	44.76	43.80	21.04	27.63
GMBT-058	08.29	40.06	55.59	16.56	38.03
GMBT-059	16.87	45.92	40.31	19.40	24.04
GMBT-060	14.28	44.92	44.23	21.78	26.89
GMBT-002	49.60	71.08	13.71	06.71	05.38
GMBT-024	24.99	53.36	30.48	14.96	16.51
GMBT-029	40.24	61.84	18.34	08.75	10.99
GMBT-014	50.36	72.43	06.32	03.43	00.10
GMBT-018	85.79	92.78	03.59	01.86	00.27

1 MPR is Matching Probability for two randomly selected individuals.

2 MPS is Matching Probability for two full-Sibs.

3 EPR is Exclusion Power when putative father is unrelated to real father.

4 EPS is Exclusion Power when putative father and real father are full-Sibs.

5 EP1 (or EPSP) is Exclusion Power when only one parent is available.

2. Genomic distribution:

We performed pair-wise linkage analysis between all markers. As it is known that at least some of these sequences are organized as minisatellite clusters (31, 35, 145), we were expecting to find tightly linked systems. We found evidence for five pairs of linked markers of which four were characterized by a recombination rate inferior to 5% (Table 3). However, two of the systems involved in a tight linkage detect non VNTR-type polymorphisms (GMBT-014, GMBT-022). The corresponding probes were probably isolated because they contain a genuine although non-polymorphic minisatellite, and were fortuitously detecting other types of polymorphism. Despite these five linked pairs, results of the linkage analysis pointed towards a scattering of these markers throughout the bovine genome.

TABLE 3

Linked Systems	θ[1]	lodscore[2]
GMBT-003 and GMBT-029	0.0%	5.00
GMBT-007 and GMBT-059	11.3%	9.11
GMBT-009 and GMBT-014	4.8%	3.74
GMBT-015 and GMBT-016	3.7%	27.00
GMBT-028 and GMBT-047	2.5%	9.40

1 θ = recombination rate.

2 pair-wise lodscores were calculated with the "LINKAGE" programs.

Reference markers for the respective synteny groups were U1:GNB1, U2:ME1, U3:NKNB, U4:MPI, U5:FOS, U6:AMY1, U7:LDHA, U8:GNB2, U9:GPI, U10:SOD1, U11:VIM, U12:GPX1, U13:MET, U14:GSR, U15:CASK, U16:ABL, U17:CRYG, U18:GGTB2, U19:CAT, U20:GLO1, U21:GH, U22:AMH,

U23:ALDH2, U24:TG, U25:CLTLA, U26:OAT, U27:DU27S1b, U28:MBP, U29:RBP3 and X:DMD. Synteny groups with highest concordancy scores, to which corresponding VNTRs were assigned, are underlined.

Evidence for a broad genomic distribution of our VNTRs was supported by the tentative assignment of 13 of them to 11 different synteny groups using somatic cell hybrids (Table 4). GMBT036 identifies a previously unmarked bovine synteny group. Probe GMBT-021 was assigned to the same synteny group as probes GMBT-015 and -016. Although the latter two probes were shown to be tightly linked, linkage between those probes and the former one could be excluded for recombination rates $\leq$ 15%.

TABLE 4

Concordany Analysis for Synteny Mapping of VNTRs

SYNTENIC GRP	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
CHROMOSOME			5	21	10				18				2		6		8
GMBT-002	65	47	55	55	45	50	45	25	69	40	75	60	70	37	45	100	70
GMBT-005	60	63	50	70	40	55	40	60	50	85	50	55	65	58	50	55	65
GMBT-006	45	42	35	65	55	70	45	65	56	80	45	40	60	58	75	40	60
GMBT-007	55	53	45	75	35	70	45	75	56	100	35	50	60	63	65	40	60
GMBT-009	45	32	55	45	55	80	65	55	63	70	55	50	50	63	65	60	50
GMBT-011	30	37	50	40	50	65	60	50	63	55	50	35	35	68	50	45	35
GMBT-015	65	63	65	95	25	60	25	65	75	70	55	60	80	47	55	60	80
GMBT-016	78	61	65	96	61	87	83	74	74	70	57	74	70	83	79	56	83
GMBT-017	70	58	70	80	40	65	40	60	63	65	60	65	80	42	60	65	96
GMBT-019	40	32	50	20	100	55	80	40	31	35	70	45	45	63	60	45	45
GMBT-021	70	68	60	100	20	55	30	60	81	75	50	65	75	53	50	55	75
GMBT-022	50	47	30	60	30	55	40	60	50	85	30	35	45	58	50	35	45
GMBT-027	55	58	55	35	65	20	45	25	44	10	75	60	40	47	25	60	40
GMBT-028	70	58	70	80	40	65	40	60	63	65	50	65	100	42	60	65	80
GMBT-036	65	58	75	55	45	50	45	45	56	50	75	70	60	42	45	70	60
GMBT-041	60	58	60	80	20	65	30	60	75	75	50	55	75	37	60	65	75

TABLE 4 (Continued)

Concordancy Analysis for Synteny Mapping of VNTRs

SYNTENIC GRP	18	19	20	21	22	23	24	25	26	27	28	29	X
CHROMOSOME		15	23	19			14		**26**		**21**		
GMBT-002	75	55	70	40	60	55	40	50	45	70	55	45	60
GMBT-005	50	40	65	75	55	80	65	67	40	65	**90**	60	25
GMBT-006	45	55	60	70	60	75	**100**	50	55	40	55	85	30
GMBT-007	45	55	70	90	70	75	80	50	55	50	75	75	10
GMBT-009	65	85	70	60	**100**	55	60	33	85	60	55	65	30
GMBT-011	60	70	55	55	85	40	55	50	**100**	45	40	60	45
GMBT-015	75	45	80	60	50	65	70	50	45	60	55	45	40
GMBT-016	88	70	83	80	80	24	72	55	80	0	54	0	30
GMBT-017	70	50	75	55	45	70	65	33	30	65	60	60	35
GMBT-019	50	70	25	35	55	50	55	50	60	45	30	60	65
GMBT-021	70	40	75	65	45	60	65	50	40	55	60	60	35
GMBT-022	30	40	55	**95**	55	60	75	67	60	35	60	70	25
GMBT-027	65	35	30	20	30	35	30	83	45	50	25	35	**100**
GMBT-028	70	50	75	55	45	70	65	33	30	65	60	60	35
GMBT-036	65	55	70	40	60	55	40	50	45	**100**	65	45	50
GMBT-041	60	50	**95**	65	65	60	65	33	50	75	70	60	25

Eight VNTRs were as well mapped by "in situ" hybridization: GMBT-006 to 14q11-16, GMBT-005 to 24q13.3-22, GMBT-011 to 26q11-21, GMBT-015 and GMBT-016 to 21q22-24, GMBT-019 to 10q14-23, GMBT-022 to 19q21-23, GMBT-028 to 2q13-21. Again good genomic coverage was evident, since six probes mapped to five different chromosomes. Probes GMBT-015 and -016 both mapped to 21q 23-24 as expected from the linkage study and the assignment on the hybrid panel. Surprisingly, five out of the eight VNTRs clearly showed an interstitial map location. Only probes GMBT-015, -016 and -022 were located proterminally, the former two identifying the same minisatellite cluster. These results seem to contrast with those of Royle et al. (35), which demonstrated preferential proterminal mapping of human VNTRs. Probe GMBT011, previously located on U26, was mapped to chromosome 26, allowing us to tentatively assign synteny group U26 to chromosome 26.

3. Conservation of sequence and map location within Bovidae:

We hybridized ten bovine VNTRs to sheep Southern blots, under slightly reduced stringency conditions. Seven of them were yielding locus specific patterns, of which six were showing a substantial degree of polymorphism in a sample of 5 unrelated sheep (Table 1).

Probes GMBT-016, -019 and -022, mapping in the bovine to 21q23-qter, 10q15-q24 and 19q21-qter respectively, were mapped by <u>in situ</u> in sheep as well. The three probes produced signals on chromosomes 18, 7 and 11, recognized as evolutionary homologues of bovine chromosomes 21, 10 and 19 (116). Moreover, the signals were found over the exact positions as expected in case of conservation of chromosomal location in cattle and sheep.

DISCUSSION

To isolate bovine VNTRs, we have used a strategy similar to Wong et al. (146, 147), based on the screening of size-selected restriction fragments obtained by complete digestion with the four-cutters MboI and HaeIII. Advantages of this strategy are: (1) the complexity of this size-range is substantially reduced; following Bishop et al. (128) and assuming an exponential distribution of restriction fragment lengths, the fragments > 2 Kb represent about 10^{-3} of the total number of MboI or HaeIII fragments, corresponding to approximately 10^4 fragments; this allows us to work readily with plasmid vectors; (2) the subsequent search for and use of the polymorphism is performed with the same enzyme used to generate the libraries, obviating the need to screen several restriction enzymes, hence reducing costs; (3) relying on frequent four-cutters, the cloned minisatellites contain very little flanking sequences and only very few of them carry highly repeated sequences which would interfere during hybridization; (4) theoretically, the larger minisatellites targeted by this approach are more likely to be involved in mutational events and could therefore be the more polymorphic ones.

A disadvantage of this approach is the unequal representation of minisatellite loci in our library. The libraries were generated with a mixture of DNA from 20 unrelated individuals, to increase the number of clonable microsatellites. As a consequence, loci for which most alleles are within the selected size range will be overrepresented, compared to loci for which the majority of alleles in the population are bellow this range.

This collection of bovine VNTRs could be used for DNA based individual identification and paternity diagnosis. Combining our top 5 probes, matching

probabilities and exclusion powers at least as good as those obtained with classical systems are obtained: MPR: $8x10^{-8}$, MPS: $4x10^{-3}$, EPR: 0.999, EPS: 0.893, EP1: 0.987. Adding more probes will of course only increase the power of the system. As a matter of fact, these probes have been used efficiently to solve paternity problems beyond the power of blood group systems. DNA typing is not limited to blood samples as present systems are, which expands its spectrum of applications and power. As an example, DNA typing has been used to deal efficiently with fetal blood cell chimerism (127), frequently encountered in cattle. Compared with multilocus DNA fingerprints, locus-specific VNTRs are much easier to interpret and are more reproducible. Following properly established standardization procedures, a "common language" could be established allowing exchange of information between laboratories. It is noteworthy that heterozygosity and allelic frequencies for some probes seem to vary substantially between breeds. As an example, probe GMBT-012 is characterized by an heterozygosity of 22% in Holsteins, but higher than 50% in both Herefords and Brown Swiss. Hence, proper use of these probes may initially require accurate estimation of genetic variation for different breeds.

Assuming a coverage of 20 cM per marker in linkage studies, the set of markers described in this paper would allow the scanning of approximately 7 Morgans. Accepting a total map length for the bovine genome of 25 Morgans (148), this represents close to 33%. We have complemented the set of bovine VNTR described in this paper with over 80 multisite haplotypes, generated with cosmid probes, and more than 100 microsatellite systems (31, 148, 149). Therefore, the majority of the bovine genome is now amenable to linkage scanning. Since several of these markers are already "anchored" to specific chromosomes or synteny groups, a primary bovine

DNA marker map should soon be available. Moreover, the remarkable conservation of mini- and microsatellites within Bovidae will substantially accelerate the construction of genetic maps in sheep and goats and offer the possibility to address interesting evolutionary issues.

EXAMPLE 2

GENERATION OF BOVINE MULTISITE HAPLOTYPES USING RANDOM COSMID CLONES.

INTRODUCTION

The possibility to generate nearly unlimited numbers of genetic markers through the study of DNA Sequence Polymorphism (DSP)(76), has revolutionized human genetics: genetic markers have been used to map genes involved in a variety of human diseases, which has direct implications for genetic counselling strategies and is a first step towards their subsequent cloning by reverse genetics; they are revolutionizing individual identification and the examination of familial relationships; and they are invaluable tools in the study of a wide variety of biological issues. In particular, they are expected to play a key role in the ongoing efforts to entirely map and sequence the human genome.

For breeders of domestic animal species, the availability of large numbers of genetic markers means, besides new approaches for individual identification and paternity diagnosis, the possibility to map and study genes determining production traits , and to use this information in marker assisted selection and velogenetics (91, 96, 141, 150).

Particularly challenging is the fact that the majority of production traits are complex,

multifactorial traits. Animal breeders, however, have the advantage that phenotypic information has been carefully recorded for thousands of animals over the years for use in classical biometrical breeding programs, and that they can, if necessary, design and generate the ideal family material required for such mapping studies.

In both the human and animal field, polymorphic markers characterized by the highest possible heterozygosities or "Polymorphism Information Content" (PIC)(76) is paramount. Hence, the focus has changed from the original diallelic Restriction Fragment Length Polymorphisms (RFLPs) to more informative systems based on the study of sequences such as minisatellites (18, 20), and more recently microsatellites (32-34) and the polydeoxyadenylate tract of SINE-repetitive elements (37). Minisatellite sequences in particular have proven very powerful. They seem to suffer, however, from a non-random genomic distribution, especially in the human where in addition, they show proterminal confinement (35). Microsatellites, although very abundant and highly polymorphic, require prior sequencing efforts to generate the primers needed for their in vitro amplification. Moreover, the large scale use of microsatellites requires the development of more efficient multiplex amplification and data collection schemes.

An alternative strategy for the generation of highly informative marker systems is to combine several, closely spaced diallelic RFLPs into more informative polyallelic multisite haplotypes (98). We have explored the use of random bovine cosmid clones in Southern blot hybridizations in order to identify such sets of closely spaced DNA Sequence Polymorphisms. Because of the population structure imposed by breeding strategies, effective population sizes of domestic species are expected to be reduced compared to humans. It was

interesting therefore, to check in how far this would decrease the observed level of genetic variation and in how far the expected concomitant increase in linkage disequilibrium would affect the efficiency of the chosen approach in domestic animal populations.

MATERIALS AND METHODS

1. Preparation of cosmid clones:

Bovine genomic DNA was prepared using standard procedures, partially digested with MboI, and size fractionated by rate zonal centrifugation in a 10%-40% sucrose gradient. Size fractions around 40 Kb were ligated into the XhoI site of the cosmid vector pWEC (pWE 15 vector (Stratagene) with pUC18 polylinker - *Erica Cumlin, personal communication*), after partial fill-in of the insert and vector sticky ends with respectively dATP, dGTP and dCTP, dTTP. The obtained constructs were packaged into Gigapack II Gold extracts (Stratagene) and used to infect E.Coli 490A hosts (gift from R. White, University of Utah, Salt Lake City, Utah, USA). 110 colonies were selected at random, cosmid DNA was prepared using standard procedures, and purified by CsCl/Ethidium Bromide isopycnic centrifugation.

2. Southern blot hybridization:

Genomic DNA from 9 unrelated Holstein individuals was prepared from venous blood using standard procedures and digested with 5U/µg of the following enzymes in the presence of 4mM spermidine: BamHI, BglI, BglII, EcoRI, EcoRV, HindIII, KpnI, MspI, PstI, PvuII, TaqI and XbaI. 4µg DNA per individual was separated according to size by agarose gel electrophoresis and blotted onto Pall Biodyne B membranes using NaOH 0.4M as transfer buffer. Membranes were prehybridized at 65 C for 4 hours in 10% PEG, 7%SDS, 50mM NaHPO4 (pH 7.2) in the presence of 350µg/ml bovine genomic DNA. Cosmid DNA was labelled by

random-priming (111) to specific activities of $.5*10^9$ cpm/μg, prehybridized with bovine genomic DNA (5mg/ml) for 90 min. at 68 C (112), and added to the prehybridized membranes for 16 hours. Final washes were in 0.1XSSC, 0.1%SDS and at 65 C. Autoradiography was carried out for 2 to 6 days at - 80 C with Kodak XAR-5 film and intensifying screens (DuPont Cronex Lightning-Plus). Membranes were stripped by boiling into 0.1%SDS and reused up to at least 10 successive times.

3. Calculation of nucleotide diversities:

Nucleotide diversities, π, corresponding to the average heterozygosity per nucleotide site were estimated following Ewens et al. (130), using:

$$\pi = \frac{\Sigma_i k}{2\Sigma r_i m_i \text{lnn}_i}$$

where n_i stands for the number of chromosomes studied with the i^{th} enzyme, r_i for the number of bp of the recognition sequence of the i^{th} enzyme, m_i for the number of cleavage sites explored with that enzyme, of which k_i are polymorphic.

The number of explored restriction sites was estimated from the number of fragments f observed by Southern blotting, using m = (3f+1)/2 (119).

RESULTS

110 randomly selected bovine cosmid clones were used in Southern blot hybridization experiments as described in Materials and Methods. 96 of them, or 87.2%, gave usable patterns and were kept for further analysis. Combining data from the 12 restriction enzymes used, a mean of 53.87 fragments per cosmid

qualified as unambiguously readable. The only RFLPs considered in this paper, are the ones affecting these selected fragments.

Assuming a cosmid insert size of 40 Kb and estimating the mean restriction fragment length in Kb (L) for a restriction enzyme according to Bishop et al. (128), the expected number of restriction fragments detected in Southern blot hybridization per cosmid probe for a given enzyme can be approximated by: integer(40/L)+1. For our 12 enzymes, we expect therefore a total of 173 fragments per cosmid probe. Therefore, the 53.87 fragments actually observed per clone, represent only 31% or less than 1/3 of what is theoretically possible. The remaining 69% are missed either because they were considered difficult to read, or more often because they went undetected due to their abundance in highly repetitive elements blocked by the competitor DNA, or due to their size, too small for efficient detection in our conditions of Southern blot hybridization. The smallest fragments qualifying as readable in this study, were in the 1 Kb size-range. The latter factor is particularly apparent with the two used four-cutters, MspI and TaqI, whose expected mean fragment length are the lowest (1747bp and 1179 bp respectively) despite the presence of the rare CpG dinucleotide. Only about 15% of the expected number of fragments are detected for these two enzymes.

Nevertheless, as much as 82 of these 96 cosmids, or 85%, were showing at least one polymorphism within our sample of 9 randomly selected individuals. The detected polymorphic events are classified into two groups: 1) Point Mutations ("PM"), whenever a defined polymorphic pattern is only seen with a single enzyme, and 2) Insertions-Deletions ("ID"), whenever such a pattern is seen with two or more enzymes. Following these rules, we identified 215 polymorphic events, or a mean of 2.6 independent RFLPs per cosmid probe. 162 of these

(75.3%) were considered of the PM type, the remaining 53 (24.7%) of the ID type. Table 5 summarizes these results.

TABLE 5

Bovine Multisite Haplotypes

NAME	BamHI	BglI	BglII	EcoRI	EcoRV
MSBT001			1ID(11) ?/3		
MSBT005	2PM(5) 6/7 3PM(5) 4.4/7.8	1ID(17) 4.2+2.9/3.9+3.2		1ID(17) 9.5+1.5/9.1+1.6	
MSBT007	1ID(17) 3/14				
MSBT009					
MSBT00A					1PM(28) ?/5.8
MSBT011	3ID(39) 7.8/5	1PM(22) 7.7/8 2PM(39) 5.5/4			
MSBT013				1ID(22) 9.5/10 2PM(6) ?/3.7	
MSBT015	1ID(44) 14/13			2PM(23) 25/22	1ID(44) 8.5/6.5
MSBT016					
MSBT017					

TABLE 5 (Continued)

NAME	HindIII	KpnI	MspI	PstI	PvuII
MSBT019		??(??)			1PM(22) 20/17+2.8 2PM(47) 30+5.5/36
MSBT020					1PM(11) 4.3/5.5
MSBT001	2ID(55) 9/12.5		1ID(11) ?/8.8 2ID(55) 12/12.5		??(??)
MSBT005			1ID(17)	1ID(17)	1ID(17) 5.8+5/5.4+4.8 4PM(50) 4/7.8
MSBT007	1ID(17) 12/6.8		1ID(17) 8.2/9.5		
MSBT009			1PM(44) 3.9/5.7		2PM(46) 4.5+5.1/9.5
MSBT00A			2PM(28) 5/6.5 3PM(11) 3.8/2.8 4PM(5) 11/9.3+1.8		
MSBT011	4PM(22) 16/14				
MSBT013		1ID(22) 19/16	??(??)		

TABLE 5 (Continued)

NAME	HindIII	KpnI	MspI	PstI	PvuII
MSBT015			??(??)		
MSBT016			1ID(40) 5.8+4.3/6.5		
MSBT017					
MSBT019		3PM(50) 11/6.4+4.3	?? (??)		
MSBT020	2PM(22) 15/10				

TABLE 5 (Continued)

NAME	TaqI	XbaI	HET.
MSBT001	??(??)		55
MSBT001			0
MSBT005	1ID(17) 4.2+2.2/3.9+2		54
MSBT005	3PM(5) 14/6		0
MSBT007			17
MSBT009			77
MSBT00A			50
MSBT00A			0
MSBT00A			0
MSBT011	3ID(39) 5.3+2/		67
MSBT011	4.4+4.6+2.9		0
MSBT013	1ID(22) 5.5/6		33
MSBT013			0
MSBT015	1ID(44) 12/9		66
MSBT015	3PM(33) 8.3/10.8		0
MSBT016	1ID(40) 6/3.3+2.5		50
MSBT017			0
MSBT019	4PM(44)		89
MSBT019	5PM(38)		0
MSBT020		3PM(35) 30/25	66

TABLE 5 (Continued)

NAME	BamHI	BglI	BglII	EcoRI	EcoRV
MSBT022	1PM(47) 3.3/3.5			2PM(47) 5.8/2.5+3.2 3PM(5) 5.8/9.2 4PM(8) ?/2.1	
MSBT023					
MSBT024				1ID(26) 14/7.5	1ID(26) 10/10.7
MSBT025		1PM(11) 9.6/7.5+2.2	2PM(33) 12/9.7		
MSBT026				1??(??) 8/?	1??(??)
MSBT028		1ID(37) 5/6.3 2ID(??) 3/?	1ID(37) 14/10 2ID(??) 15/?	1ID(37) 13/6.6	1ID(37) 8.1/5.5 2ID(??) 3.4+5/?
MSBT030	1PM(67) 8/5.8+2				
MSBT031					??(??)
MSBT032	1PM(22) 9.6+6/16			2ID(11) ?/6.2+7+12	2ID(11) ?/6.1
MSBT033				1PM(11) 3.5/5 2ID(44) 5.8/7 3PM(44) 9/12	2ID(44) 6/7.8 ??(??)
MSBT034					

TABLE 5 (Continued)

NAME	HindIII	KpnI	MspI	PstI	PvuII
MSBT022					
MSBT023					1PM(45) 5.5+3.7/3.2+2
MSBT024	1ID(26) ?/7.8		1ID(26) 7.5/5		1ID(26) 13/14
MSBT025	??(??)				
MSBT026	?		2??(??)		
MSBT028					
MSBT030					3??(??) 3.3/3.1
MSBT031	??(??)		??(??)		??(??)
MSBT032	2ID(11) ?/12 3PM(33) 4.6/4.9 4ID(56) 4/3.4				2ID(11) ?/2.7
MSBT033	2ID(44) 14/15 ??(??)	4ID(66) 6.5/2.5	4ID(66) 1.7/4.3		??(??)
MSBT034		1ID(44) 5.9/3.9	??(??)		1ID(44) 4.7/4.3 2PM(22) 5.3/4

TABLE 5 (Continued)

NAME	TaqI	XbaI	HET.
MSBT022	5PM(44) 9/3.2		89
MSBT022	??(??)		0
MSBT022			0
MSBT023		2PM(33) 17/13	45
MSBT024	1ID(26) 14/15	1ID(26) 3.8/4.7	26
MSBT024	??(??)		0
MSBT025	3PM(37) 9/11/10		78
MSBT025	4PM(??)		0
MSBT026			0
MSBT028			37
MSBT02E			0
MSBT030	2PM(11) 5/5.2		78
MSBT03:		??(??)	0
MSBT032	4ID(56) 4.1/4.3		67
MSBT032			0
MSBT032			0
MSBT033	5PM(44) 4/3.4	2ID(44) 5/6.9	88
MSBT033		??(??)	0
MSBT033			0
MSBT034	1ID(44) 5.8/8		56
MSBT034	3??(??)		0

TABLE 5 (Continued)

NAME	BamHI	BglI	BglII	EcoRI	EcoRV
MSBT035			1ID(55) 6.3/6.6		
MSBT037			3??(11) 2??(11)		
MSBT038	5PM(??)	1ID(11) 6.8/10.5	1ID(11) 15.0/14.0	1ID(11) 7.5/6.4	1ID(11) 7.5/6.4
MSBT039	1PM(11) 9/9.6				
MSBT040			1PM(22) 18/16		
MSBT041					
MSBT042	1PM(33) 14/15 2PM(33) 6/2				
MSBT043		??(??)	1ID(44) 4.5/4.7 2ID(55) 3.7/4.3	1ID(44) 3.2/3.6	
MSBT044					
MSBT045			1ID(44) 4/3.4	2ID(89) 3.9/6.5/3.5	2ID(89) 4/4.8/3.5
MSBT046				1ID(78) 3.5/3.6	1ID(78) 3.5/3.6
MSBT047	1PM(11) 6.3/7.3 2PM(30) ?/10.5				

TABLE 5 (Continued)

NAME	HindIII	KpnI	MspI	PstI	PvuII
MSBT035			2ID(??)		2ID(??)
MSBT037				1PM(22) 4.5/5 4??(??) 3.8/?	
MSBT038	1ID(11) 9.6/8.5	1ID(11) 6.0/5.5	2PM(37) 19/2.7+16 4??(??) 4.3/?		1ID(11) 6+8/2
MSBT039					
MSBT040			2??(??)		
MSBT041					
MSBT042		3PM(44) 7/?	??(??)		
MSBT043	1ID(44) 6.5/6.9 3PM(78) 4.8/3.3 4PM(78) 4.8/4.6 9PM(??) 3/2		5PM(33) 2.9/3.1 6ID(33) 3.7/3.9		6ID(33) ?/2+2.1 7PM(11) ?/2.4
MSBT044	0				1ID(??) 3.85/3.9/3.95..
MSBT045	1ID(44) 11/10.5		1ID(44)		
MSBT046					
MSBT047			3PM(11) 6.5+2.8/9.3 4PM(??) 19/14		

TABLE 5 (Continued)

NAME	TaqI	XbaI	HET.
MSBT035	1ID(55) 7.8/5.3		55
MSBT037			22
MSBT037			0
MSBT038	1ID(11) 5.5/?	1ID(11) 6.6/5.6	44
MSBT038	3PM(22) 3.5/?	6??(??)	0
MSBT039			11
MSBT040			22
MSBT041			0
MSBT042	??(??)	??(??)	44
MSBT042			0
MSBT043	8PM(11) 3.3/2.9		99
MSBT043	2ID(55) 1.4/2.2		0
MSBT043			0
MSBT043			0
MSBT044			0
MSBT045	3PM(33) 11/12		89
MSBT046	2PM(33) 4.9/5.2		78
MSBT046	1ID(78) 2.9/1.7		0
MSBT047	5PM(??) 14/21		44
MSBT047			0

TABLE 5 (Continued)

NAME	BamHI	BglI	BglII	EcoRI	EcoRV
MSBT048					
MSBT049				1ID(22) 9/3.9+5.3	1ID(22) 9.7/4
MSBT050				1PM(??) ?/8.2	
MSBT051	2PM(22) 20/18		1PM(11) 7/14		
MSBT052			1ID(11) 17/14		2ID(??)
MSBT053					1PM(??)
MSBT054					
MSBT056			1ID(55)	1ID(55) 5.6/7 2PM(22) 12/5.5	1ID(55) 5.7/7.3
MSBT057				1ID(11) ?/8	1ID(11) ?/8.5 2PM(??) 7/6.7
MSBT059				1ID(33) 1.8/3.4	1ID(33) ?/3.2
MSBT060		1ID(11) 6/5.8	1ID(11) 5/6.5	1ID(11) 9.6/11 2ID(55) 6.4/9.6	1ID(11) 9.6/11.5 2ID(55) 3.5/6.6 3PM(22) 9.6/8.8
MSBT061					

TABLE 5 (Continued)

NAME	HindIII	KpnI	MspI	PstI	PvuII
MSBT048	1ID(44) 6.2/5.9+3	1ID(44) 27/15+7.8	1ID(44) 4.3/3.7+3.2 2PM(11) 5.5/4.6		1ID(44) 6.5/9
MSBT049	2PM(11) 8.4/7		3PM(??)		
MSBT050	2PM(??) ?/1.5	2ID(33)		3PM(11) 3.2/2.8	2ID(33) 2+1.6/6.8+1.8
MSBT051			3PM(??)		
MSBT052			1ID(??)		1ID(11) 12/10.5
MSBT053				2ID(44) ?/6.8	2ID(44) 2.1/2.4 3PM(??) 11.5/11
MSBT054				1PM(22) 2.6/2.9	
MSBT056					1ID(55) 3/2.7
MSBT057					
MSBT059			1ID(33) 12/11		
MSBT060	1ID(11) 6.2/6.8			4PM(??) 2.7/1.4	
MSBT061	1ID(44) 7+2.4/9.3				1ID(44) 4.7/5.2

TABLE 5 (Continued)

NAME	TaqI	XbaI	HET.
MSBT048		1ID(44) 6.4/7	55
MSBT048			0
MSBT049			22
MSBT050	2ID(33) 2.5+2.7/2.4+2.9		55
MSBT051	4PM(22) 5.2/6		78
MSBT051	5PM(33) 3.3/3.1		0
MSBT052	1ID(11) 8/7.8		11
MSBT053	??		44
MSBT053			0
MSBT054	2PM(11) 4.6/7.4	3PM(33) 3.7/4.8	33
MSBT056	1ID(55) 2.5/4.7		66
MSBT056			0
MSBT057	3PM(11) 14/9.5	4PM(44) 20/15.5+4.5	66
MSBT057			0
MSBT059	2PM(66) 7.8/6.8		66
MSBT059	3PM(??)		0
MSBT060			66
MSBT060			0
MSBT060			0
MSBT061			44

TABLE 5 (Continued)

NAME	BamHI	BglI	BglII	EcoRI	EcoRV
MSBT062					??(??)
MSBT064					
MSBT065			1ID(44) 4.5/7		
MSBT067					1ID(55) 9/7.3
MSBT068	1ID(11) 18/11			1ID(11) 14/12	1ID(11) 10/8.8
MSBT069	2PM(33) 16+3.5/19.5		1PM(44) 12.5/10.5	3PM(11) ?/9	4PM(22) 9/8.4
MSBT070		1PM(11) ?/3		2ID(55) 2.5/2.9	2ID(55) 2.7/3.2
MSBT071					
MSBT072			1ID(33) 13/12		
MSBT074					
MSBT075					
MSBT076	2ID(44) 6.2/4.8		1ID(33) 2.5/2	2ID(44) 2/2.8	2ID(44) 2.3/2.9
MSBT078				1PM(11) 7.6/6.7	2PM(11) 7.8/6.7
MSBT079	??(??)				
MSBT080			1ID(11) 4.5+2/5.7+4.3		
MSBT081			??(??)		??(??) ?/11

TABLE 5 (Continued)

NAME	HindIII	KpnI	MspI	PstI	PvuII
MSBT062					
MSBT064		1PM(33) 25/19			
MSBT065			1ID(44) 5.2/4		
MSBT067			1ID(55) 3.9/3.5		
MSBT068					
MSBT069					5PM(33) ?/3.2
MSBT070	3PM(22) 9.4/7.9+1.5	4PM(11) 13/10.7+2			5PM(44) 8.7/9.5
MSBT071					1PM(??) ?/8
MSBT072	2PM(11) 9.5/8.2		1ID(33) 3.7/4	3PM(??) ?/5.4	4PM(44) 5.9/9
MSBT074	1PM(22) 4.6/5.3		2PM(44) 10.8/12/9.5 3PM(33) 17/16.5 4PM(44) 10.8/12/9.5		5PM(33) 3.3/5.7
MSBT075					
MSBT076					
MSBT078			3PM(11) ?/6		
MSBT079					
MSBT080			1ID(11) 2/2.6+2.7		1ID(11) 11/8+2.8
MSBT081			2PM(33) 2.2/2.7		

TABLE 5 (Continued)

NAME	TaqI	XbaI	HET.
MSBT062	??(??)		0
MSBT064			33
MSBT065	1ID(44) 6.7/4.7	1ID(44) 4.8/3.1	44
MSBT067		2PM(??)	55
MSBT068	2PM(33) 3/1.6		44
MSBT069			89
MSBT070	6PM(??) ?/3.7		78
MSBT071	2PM(11) 15/11	3PM(??)	11
MSBT072		5PM(11) 20/12	78
MSBT074			66
MSBT074			0
MSBT074			0
MSBT075	1PM(11) 3.2/4.4		11
MSBT076	1ID(33) 2/3.8		44
MSBT078		4PM(33) 4/6.4	44
MSBT079		1PM(11) 11/13	44
MSBT079		2PM(33) 3.5/3.7	0
MSBT080		2PM(11) 6.3/3.9+2.5	22
MSBT081			33

TABLE 5 (Continued)

NAME	BamHI	BglI	BglII	EcoRI	EcoRV
MSBT083	2PM(44) 11/17	1PM(11) ?/4.5		3ID(33) 5/3.8	3ID(33) 5.5/3.9
MSBT084	1PM(44) 4.5/9				
MSBT085		1ID(44) 7.7/6.5 2ID(22) 3.3/4.8			3ID(55) 5.5/4.8 2ID(22) 6/4.5
MSBT086	2ID(22) 10.5+2.2/12.5		1PM(11) 6.2/6+9.4		3PM(11) ?/3.7
MSBT087					
MSBT088					
MSBT089					
MSBT090					
MSBT091					??(??) 5.5/6
MSBT092					
MSBT093					
MSBT094					
MSBT096					
MSBT097					
MSBT098					

TABLE 5 (Continued)

NAME	HindIII	KpnI	MspI	PstI	PvuII
MSBT083	4PM(??) ?/7				
MSBT084				2PM(22) 5.1/3.9	
MSBT085	2ID(22) 12/5.7				
MSBT086			2ID(22) 18/21 4ID(55) 4.3/7	4ID(55) 2.5/5	5PM(55) 7.1/5
MSBT087					
MSBT088					
MSBT089					
MSBT090					
MSBT091			??(??)	3ID(33) ?/4	
MSBT092	1PM(44) 4.3/2.8		2PM(22) 3.5/4		
MSBT093				1PM(33) 4.7/1.9+2.5 2PM(33) ?/3.3	
MSBT094					
MSBT096		1PM(11) 8/4.8+3.2			
MSBT097	1PM(??)				
MSBT098				1PM(66)	2PM(??)

TABLE 5 (Continued)

NAME	TaqI	XbaI	HET.
MSBT083	5PM(11) ?/4		66
MSBT084	3PM(33) 2.3/2.4		55
MSBT084	4PM(??) ?/2.8		0
MSBT085	3ID(55) 5/3.8	2ID(22) 11+15/23	55
MSBT085	1ID(44) ?/2.7		0
MSBT086	4ID(55) 6/3.3		78
MSBT086			0
MSBT087			0
MSBT088			0
MSBT089			0
MSBT090	1PM(11) 1.8/3.5		11
MSBT091	3ID(33) ?/5.8		33
MSBT092	3PM(33) 5.3/6.8		66
MSBT093			44
MSBT093			0
MSBT094			0
MSBT096			11
MSBT097		2PM(??)	0
MSBT098	3PM(55) 4.5/4	5PM(11) 18/15	66

TABLE 5 (Continued)

NAME	BamHI	BglI	BglII	EcoRI	EcoRV
MSBT098					
MSBT099	1PM(33) 14/15			2ID(11) 10.5+4/15+3.9	2ID(11) 8+4/8.8+3.9
MSBT100					
MSBT101	2PM(22) 8/10.2	1PM(44) 9+7/16			
MSBT103					
MSBT104					
MSBT105				1PM(??)	
MSBT106	1ID(22) ?/9.5			1ID(22) ?/8.3	1ID(22) ?/8.5
MSBT107	2PM(67) 12+2.4/14.4	1ID(33) 2.6+2.8/3.1			
MSBT108	2PM(11) 9.5/7.8		1PM(11) 15/6.5		
MSBT109					
MSBT110					1PM(44) 10.3/9.1
MSBT111				1PM(11) ?/3	
MSBT113				1PM(22) 10/9.5	
MSBT114	1ID(33) ?/4.4			1ID(33) 8.8/8+1.9	1ID(33) 5.5/4.8

TABLE 5 (Continued)

NAME	HindIII	KpnI	MspI	PstI	PvuII
MSBT098					
MSBT099		2ID(11) 19/21 3PM(78) 10+9/19	2ID(11) 2.1/1.9	2ID(11) 4.9/4.8 4PM(??) ?/4.2	
MSBT100		1PM(11) 8/11.5			
MSBT101			??(??)		3PM(55) 3.55/3.6
MSBT103					
MSBT104					
MSBT105			2PM(??)		
MSBT106	1ID(22) ?/3.7				
MSBT107			3??(??)		
MSBT108					
MSBT109	1ID(55) 4.9+5.8/10.8			1ID(55) 4.6/3.4	
MSBT110					
MSBT111				2ID(22) 2.9/3.2/2.8	3ID(11) 7.5/9.2
MSBT113	2PM(33) 7.5/3.4+4.2 3ID(11) 4/7+3.8	3ID(11) ?/10 4ID(55) 12+21/35			
MSBT114	1ID(33) ?/2.5		1ID(33) 11/10	1ID(33) 5.3/5.9+1.6	1ID(33) ?/4

TABLE 5 (Continued)

NAME	TaqI	XbaI	HET.
MSBT098	4PM(11) 2.5/2.8		0
MSBT099	2ID(11) 4.9/4.8		78
MSBT099	5PM(22) 2.5/1.5		0
MSBT100			11
MSBT101		4PM(33) 18/5/7.3	66
MSBT101		5PM(33) 18/5/7.3	0
MSBT103			0
MSBT104			0
MSBT105			0
MSBT106			22
MSBT107	1ID(33) 2.1/2.6		78
MSBT108		3PM(50) 7/11.5	55
MSBT109		2PM(??)	55
MSBT110		2PM(??)	44
MSBT111	3ID(11) 7/8.8		22
MSBT111	4PM(11) 3.7/4.9		0
MSBT113	3ID(11) ?/7		66
MSBT113	4ID(55) 6.5/6.9		66
MSBT114		1ID(33) ?/4.9	55

TABLE 5 (Continued)

NAME	BamHI	BglI	BglII	EcoRI	EcoRV
MSBT114				2PM(22) ?/3.5	
MSBT116					
MSBT119					
MSBT120					1PM(11) ?/7
MSBT121				1PM(33) 5+5.8/10.8	2PM(33) 5/5.4

TABLE 5 (Continued)

NAME	HindIII	KpnI	MspI	PstI	PvuII
MSBT114	3PM(22) 8.7/10.8				4PM(55) 5/4.7/7.2 5PM(55) 5/4.7/7.2
MSBT116			??(??)		1PM(33) 6.5/5.5
MSBT119	1ID(55) 11/5.2				
MSBT120	2PM(44) 7.5/6.1				
MSBT121					

TABLE 5 (Continued)

NAME	TaqI	XbaI	HET.
MSBT114			0
MSBT114			0
MSBT116			33
MSBT119	1ID(55) 4.2/2.8		67
MSBT119	2PM(44) 6/7.7		0
MSBT120	3PM(44) 5.2/3.7		55
MSBT121		3PM(44) 3.4/5.2	67

Although this way of classifying RFLPs gives a conservative estimate of the number of identified polymorphisms, for cosmids characterized by strong linkage disequilibrium, the number of ID events may be inflated at the expense of the actual number of PMs. To compensate for this, we performed a second set of calculations in which a polymorphic event must be detected with at least 3 enzymes to qualify as ID. RFLPs previously attributed to insertion-deletion events because detected with two enzymes, are now considered as two independent point mutations. 24 polymorphisms initially considered IDs fell into this category. Following this approach, 239 independent RFLPs were identified or 2.9 per cosmid, with now 87.9% of the PM type and 12.1% of the ID type.

Table 5 reports the observed heterozygosities obtained with the generated multisite haplotype systems. These values correspond to the percentage individuals heterozygous for at least one of the polymorphisms identified with a given cosmid. Noteworthy, this parameter is not affected by the mode of classification of RFLPs in PMs or IDs. At this point and without segregation information, we can't dissect the heterozygous genotypes into their component haplotypes. These heterozygosities were estimated on a small sample and should therefore be considered cautiously. As pointed out by Skolnick and White (129), the main advantage of working with a sample of 9 individuals is that it is sufficient to identify the majority of useful polymorphisms. However, the mean heterozygosity of 51.9%, obtained over the 84 polymorphic cosmids demonstrates the power of the approach.

The following numbers of RFLPs were detected by each enzyme, irrespective of PM or ID type: TaqI: 57, EcoRI: 37, MspI: 33, HindIII: 31, PvuII: 30, EcoRV: 29, BamHI: 26, BglII: 24, XbaI: 21, PstI: 18, KpnI: 16 and BglI: 13. The number of polymorphisms detected with

BglII and PstI were corrected to adjust for the lower number of probes used with these enzymes.

Proper interpretation of the polymorphic patterns has been confirmed by segregation analysis in pedigree material for most of the described RFLPs (see hereafter).

RFLPs of the PM type were used to calculate nucleotide diversities as described in Materials and Methods. Two sets of values are reported, depending on which of the two criteria were used to classify an RFLP into the PM or ID type. Global nucleotide diversities of respectively 0.000652 and 0.000846 were obtained, meaning that a randomly selected Holstein animal will be heterozygous for approximately 1 every 1200 to 1500 base pairs. As expected because of the presence of an hypermutable cytosine followed by guanine in their recognition sequence, nucleotide diversities more than twice as high are obtained when combining data obtained with the enzymes MspI and TaqI: 0.001493 and 0.002239 respectively (5). On the other hand, the recognition sequence of the enzymes BglII, HindIII, PstI, PvuII and XbaI are devoid of hypermutable cytosines in the CpG dinucleotide and yield combined nucleotide diversities less than one third the values found with MspI and TaqI (0.000492 and 0.000648). Using these two sets of values, one can extrapolate what nucleotide diversities would be obtained if sampling hypothetical sequences composed entirely of hypermutable cytosines, giving respectively 0.004496 and 0.007012. Assuming that the majority of detected polymorphisms behave according to the neutral mutation-random drift hypothesis, nucleotide diversity and mutation rate are simply related as:

$$\pi \approx 4Ne\mu,$$

where Ne stands for the effective population size and μ for the mutation rate (1). Therefore, our data allow us to estimate that cytosines followed by guanines mutate at a rate approximately 10 times higher compared to

other nucleotides, presumably because a substantial fraction of these are methylated in the germline and prone to mutate to thymidine by spontaneous deamination.

DISCUSSION

We demonstrate in this work that large numbers of DNA markers with very acceptable Polymorphism Information Content can be quickly generated using large, randomly selected genomic probes in Southern blot hybridization experiments. The multisite haplotypes identified in this study using cosmid probes have a mean heterozygosity of 51.9% This value is of the same order of magnitude as the heterozygosities that we have obtained with a panel of approximately 40 bovine Variable Number of Tandem Repeat markers (mean heterozygosity 59%; 150), and with more than 50 bovine (TG)-dinucleotide microsatellites (mean heterozygosity 56%; unpublished).

A remarkably high proportion of the tested cosmids proved informative: 74.5% of all tested clones, and as high as 85% when considering only the clones giving readable patterns. Compared to strategies aimed at isolating hypermutable sequences such as mini- or microsatellites, very little time and effort is wasted into candidate clones which have to be dropped at a later stage.

Because the cosmid clones used as probes are selected at random, we can reasonably assume that the coverage obtained with the generated markers is fairly uniform. Monte-Carlo simulations allow us to estimate that these 82 markers are covering 29%, 47% or 60% of the bovine genome in linkage studies if a maximum of respectively 5, 10 and 15 cM are scanned on each side of each marker, and assuming a total bovine map length of 25 Morgan as deduced from chiasmata counts (151).

The Southern blot hybridization procedure used for the detection of these RFLPs is a very mature and robust methodology, allowing the treatment of very large numbers of samples simultaneously, and benefitting from intrinsic "multiplexing" properties especially when using nylon membranes. Indeed, and despite variations between batches, we are routinely using membranes for 10 or more hybridization cycles.

The main disadvantage of multisite haplotypes is the requirement to use several restriction enzymes to fully exploit their PIC. This increases costs, amounts of required DNA, complicates the organization of genotype collection and their subsequent use in linkage analysis.

The fact that 75 to 85% of the cosmids tested in cattle reveal polymorphism, compares favorably with results previously reported in the human. Schumm et al. (131) for instance report that 30% of the 1664 lambda clones they tested in a sample of 5 individuals, gave polymorphic patterns. Adjusting for the sample size and a ratio of approximately 2.5 between cosmid and phage insert size, the two figures are probably fairly similar. Surprisingly, only 54 of 101 human cosmid clones tested by the same group (152) were revealing RFLPs when tested with 9 restriction enzymes, versus 74.5% in our study with, however, 12 enzymes. It has been speculated that the relatively low level of polymorphism found in this study might result from a biais against human methylated sequences (including CpG present in the recognition sequence of TaqI and MspI) when construction the cosmid library, due to the active modified cytosine restriction system (mcr) of the E.Coli 1046 host (152).

These results are quite unexpected. Indeed, because of the population structure imposed by breeding strategies, the effective population size, Ne, in cattle is expected to be significantly lower than in the human.

In American Holstein, Ne is estimated between 10^2 and $2x10^3$ (Ina Hoeschele, personal communication). This value has to be compared with an estimated Ne of 10^4 in the human (18). Assuming identical mutation rates, this reduction of Ne should be accompanied by a concomitant reduction in genetic variation. As a matter of fact, we obtain estimates of global nucleotide diversity between 0.000652 and 0.000846 which are between 3.5 and 2.5 times lower than values typically found in human populations (2, 119, 132). This confirms our previous results in another cattle population: the Belgian Blue Cattle breed (3, 4).

At least part of the discrepancy between an apparently reduced nucleotide diversity but similar RFLP frequency, may be accounted for by the apparently higher frequency of insertion-deletion events found in cattle compared to human. Schumm et al. (131) report that 58 out of the 515 polymorphic loci (11.26%), show insertion-deletion type RFLPs; other groups report even lower frequencies of such events in the human (R.White, personal communication).. In cattle, we found that 29 (35%) to 46 (56%) out of 82 polymorphic cosmids show such insertion-deletion events, depending on whether an ID-type polymorphism has to be detected with two or more enzymes. These results seems to point towards a fundamentally different property of both genomes. It is tempting to speculate that this high level of insertion-deletions in the bovine genome reflect the activity of a mobile element. Analysis of the restriction patterns characterizing these ID events, however, does not reveal any typical, recurrent "signature" of such an element.

Altogether, our laboratory has now isolated more than 200 DNA markers for cattle with a mean heterozygosity above 50%: 82 multisite haplotypes, 40 Variable Number of Tandem Repeat markers (150) and more than 80 dinucleotide microsatellites (unpublished). The coverage of the bovine genome obtained with increasing

number of randomly selected probes (0 to 500), was estimated by Monte-Carlo simulation assuming that the used family material is sufficent to detect linkage at respectively 5, 10 and 15 cM, andg a total bovine map length of 25 Morgan as deduced from chiasmata counts (151), divided over the 30 bovine chromosomes according to their relative length. With the PIC characterizing our marker set, we feel fairly comfortable that in the majority of situations we will be able to cover genetic distances of the order of 10 cM or more, especially if applying multilocus or interval mapping techniques (66, 68). Therefore our panel of probes should cover around 75% of the bovine genome.

It is obvious that we are approaching a point where the efficiency of a strategy based on the further accumulation of random markers become questionable. After 200 probes, the additional coverage obtained per new marker, expressed as a fraction of the maximum coverage possible, now is approximately 1/5 of the coverage obtained when we started this project. This creates the need for more targeted approaches. In this regard, mappers of domestic animal genomes will benefit from the human mapping efforts and the remarkable chromosomal conservation observed within mammals (133). Based on comparative mapping information, it should be possible to identify genes likely located in the "holes" left by the random approach and to generate multisite haplotypes or microsatellite markers around their bovine homologues. We are presently exploring the feasibility of such approaches.

With the markers available today, a substantial part of the bovine genome is now amenable to linkage scanning, which will hopefully allow the mapping of Economic Trait Loci and testing the feasibility of Marker Assisted Selection schemes.

EXAMPLE 3

CLONING, CARACTERIZATION AND "IN VITRO" AMPLIFICATION OF BOVINE MICROSATELLITES

INTRODUCTION

Recently, microsatellites were proven to be an abundant source of highly polymorphic markers in the human (32-34). As their name implies, microsatellites are minute VNTR markers (18-20), characterized by tandem repetitions of very short repeats, one to four base pairs in length. Microsatellites exhibit levels of polymorphism comparable to VNTRs, but are much more abundant and apparently evenly spread throughout the genome.

We describe the cloning, caracterization and "in vitro" amplification of more than 100 such bovine microsatellites.

MATERIALS AND METHODS

1. DNA Database Search

Bovine and ovine sequences in the EMBL and Genbank (version 64.0) were searched for all types of dinucleotide and trinucleotide repeats using the Intelligenetics software, release 5.37. The minimum number of repeats was set at six. Six bovine sequences, characterized by the longest microsatellites, were retained for further analysis and are listed in Table 6.

TABLE 6

GBCYP21	GBIRBP
GBFSH	GBKCAS
GBGAPR	GBPRLGR

2. Isolation of Bovine Microsatellites

Bovine genomic DNA was digested to completion with MboI and size-fractionated by agarose-gel electrophoresis. Fragments between 250 and 500 base pairs were recovered and purified using "Gene-Clean", ligated into the BAP-dephosphorylated BamHI site of pUC13 (Pharmacia), and cloned into E. Coli DH5α cells (BRL). The resulting clones were screened for the presence of $(TG)_n$ microsatellites using a ^{32}P kinased $(AC)_{15}$ oligonucleotide as probe, and for (AG)n microsatellites using a (TC)15 probe. The library was made with female DNA to avoid the previously characterized Y-specific TG-rich bovine DYZ1 sequence (117).

3. Sequencing of Bovine Microsatellites

Positive clones were sequenced using one of the following procedures:

(a) Plasmid DNA was prepared using standard "boiling mini-prep" procedures and subjected to two chain-termination sequencing reactions using unmodified T7 DNA Polymerase (Pharmacia), with the "universal" and "reverse" sequencing primers, respectively. The ^{35}S labeled sequencing products were analyzed on standard denaturing polyacrylamide sequencing gels and detected by autoradiography.

(b) Magnetic solid-phase sequencing (137). Alternatively, positive colonies were grown in microtiter-format using standard procedures in order to establish glycerol stocks and 5μl of culture directly subjected to PCR amplifications using the following vector-specific primers:

UNIBIS: 5'-GATGTGCTGCAAGGCGATTAAGTTG-3'
REVBIS: 5'-CGGCTCGTATGTTGTGTGGAATTGT-3'

Two 30 cycle amplifications were carried out per clone, one with the UNIBIS primer biotinylated, the other with

the REVBIS primer biotinylated. Denaturation was at 93°C for 1 min. (except for the first cycle: 95°C for 5 min.) annealing at 60°C for 2 min., and extension at 72°C for 2 min. All of the PCR reactions were performed in micro-titer-format using the TECHNE MW-2 heating device. The biotinilated strand of the PCR-product was captured using the DYNAL streptavidin-coated magnetic beads according to the manufacturer recommended conditions and sequenced using unmodified T7 DNA polymerase (Pharmacia) as specified by the manufacturer.

4. Amplification and Detection of Bovine Microsatellites

(a) Simplex Amplification. The generated sequences are organized in the following way:

5'-....(UP).....$(TG)_n$.....-3'
3'-.....$(AC)_n$.....(DN)....-5'

Suitable primers for in vitro amplification are identified in "UPSTREAM (UP)" and "DOWNSTREAM (DN)" strands using the "OPTIPRIM" program designed by one of us.

Given sequence information flanking a DNA stretch, "Optiprim" helps the user to identify suitable primer pairs for PCR amplification of the corresponding DNA stretch.

Description of the program: The two DNA sequences flanking the DNA stretch to be amplified are referred to as the upstream (UP) and downstream (DN) sequence, respectively. Both for UP and DN, Optiprim tests all possible primers of given length (as defined by the user) and retains the primers (1) with melting temperature (Tm) within the range defined by the user (Tm is calculated as 2C x number of As, or Ts + 4C x number of Gs or Cs), (2) with a minimum percentage of each nucleotide as defined by the user, and (3) which cannot form secondary bonds that can be formed between two molecules

of the defined primer when sliding them in antiparallel orientation against each other, as illustrated in the following:

5'-PRIMER-3' --->

<--- 3'-REMIRP-5'

An A facing a T contributes two hydrogen bonds, and a G facing a C contributes three hydrogen bonds. No loop formation is considered when performing this analysis. This generates two sets of selected primers: an UP set and a DN set. All possible pairs of one UP and one Dn primer are then tested. Optiprim retains the primer pairs if (1) the difference between melting temperatures of the two primers is within a range defined by the user, (2) the two primers cannot form secondary structures, determined as for individual primers, except that now the UP primer is slided versus the DN primer. Using this program, 80% of the selected primer pairs were giving successful PCR amplification in our microsatellite systems. The following criteria are considered by "OPTIPRIM" when searching for primers: primer length, melting temperature and secondary structures that can be formed within and between primers. The selected primers are synthesized by phosphoramidite chemistry on Applied Biosystem synthesizers and used without further purification. The microsatellites are amplified <u>in vitro</u>, in microtiter plates and using the Techne MW2 device, in the following conditions (typically, 30µl reactions):

Target DNA	50 ng-100ng
KCl	50mM
Tris-HCl, pH 8.4	10mM
MgCl2	1.5mM
Gelatine	0.01%
dNTP	200µM each
Primers	1µM each
$dCTP^{32}$	2µCi/30µl
Amplitaq	1U/30µl

Thirty cycle amplifications are performed, characterized by a 93°C denaturation for 1 min. (except for the first cycle: 95°C, 5 min.), annealing at 55°C, 60°C or 65°C for 2 min. depending on the primers, and extension at 72°C for 2 min. Annealing temperatures are reduced by 5°C when using bovine primers on ovine target DNA.

(b) Multiplex Amplification. When performing multiplex amplifications, concentrations of KCL, Tris-HCl, MgCl2, gelatine and dNTPs are increased by 50%, while the primer concentrations are decreased to 160μM each.

(c) Detection of Microsatellite Products. 2μl of PCR reaction are mixed with the same volume of formamide dye and run in a denaturing 7% acrylamide, 32% formamide, 5.6 M urea, 13.5 mM Tris, 4.5 mM Boric Acid, 250 μM EDTA gel. ^{32}P labeled products are detected by autoradiography.

RESULTS AND DISCUSSION

1. Cloning and sequence caracterization of bovine microsatellites:

A library of MboI fragments between 250 and 500 bp was screened with the oligonucleotide probes. One out of 50 clones cross-hybridized. Assuming independent distribution of microsatellites and MboI sites, the frequency of $(TG)_{n\geq 9}$ microsatellites in the bovine genome is estimated to be at ≥100,000. Table 7 summarizes the sequence information available for about 230 such bovine microsatellites. For each of these, sufficient sequence information has been gathered to generate the required primers for PCR amplification of the corresponding microsatellite. All sequences were generated by sequencing as described above or by screening GENEBANK.

TABLE 7

Bovine Microsatellites

Sequence Name	Sequence Identification Numbers Up	Repeat	Down
AGLA13	1	2	3
AGLA17	4	5	6
AGLA206	7	8	9
AGLA209	10	11	12
AGLA212	13	14	15
AGLA215	16	17	18
AGLA217	19	20	21
AGLA218	22	23	24
AGLA22	25	26	27
AGLA220	28	29	30
AGLA223	31	32	33
AGLA226	34	35	36
AGLA227B	37	38	39
AGLA230	40	41	42
AGLA232	43	44	45
AGLA233	46	47	48
AGLA234	49	50	51
AGLA243	52	53	54
AGLA247	55	56	57
AGLA248	58	59	60
AGLA254	61	62	63
AGLA255	64	65	66
AGLA257	67	68	69
AGLA258	70	71	72
AGLA259	73	74	75
AGLA260	76	77	78
AGLA267	79	80	81
AGLA269	82	83	84
AGLA272	85	86	87
AGLA280	88	89	90
AGLA284	91	92	93
AGLA285	94	95	96
AGLA29	97	98	99
AGLA291	100	101	102
AGLA293	103	104	105
AGLA296	106	107	108
AGLA298	109	110	111
AGLA299	112	113	114
AGLA300	115	116	117
AGLA33	118	119	120
AGLA8	121	122	123

TABLE 7 (Continued)

GBFSH	124	125	126
GBIRBP	127	128	129
GBKCAS	130	131	132
GBPRLGR	133	134	135
MGTG1	136	137	138
MGTG11	139	140	141
MGTG13A	142	143	144
MGTG13B	145	146	147
MGTG3	148	149	150
MGTG4B	151	152	153
MGTG7	154	155	156
TGLA10	157	158	159
TGLA102	160	161	162
TGLA109	163	164	165
TGLA110	166	167	168
TGLA111	169	170	171
TGLA112	172	173	174
TGLA116	175	176	177
TGLA117	178	179	180
TGLA12	181	182	183
TGLA122	184	185	186
TGLA123	187	188	189
TGLA124	190	191	192
TGLA125	193	194	195
TGLA126	196	197	198
TGLA127	199	200	201
TGLA128	202	203	204
TGLA13	205	206	207
TGLA130	208	209	210
TGLA131	211	212	213
TGLA132	214	215	216
TGLA134	217	218	219
TGLA135	220	221	222
TGLA137	223	224	225
TGLA141	226	227	228
TGLA142	229	230	231
TGLA147	232	233	234
TGLA149	235	236	237
TGLA15	238	239	240
TGLA153	241	242	243

TABLE 7 (Continued)

TGLA154	244	245	246
TGLA158	247	248	249
TGLA159	250	251	252
TGLA160	253	254	255
TGLA162	256	257	258
TGLA164	259	260	261
TGLA17	262	263	264
TGLA170	265	266	267
TGLA171	268	269	270
TGLA172	271	272	273
TGLA175	274	275	276
TGLA176	277	278	279
TGLA179	280	281	282
TGLA182	283	284	285
TGLA188	286	287	288
TGLA189	289	290	291
TGLA2	292	293	294
TGLA20	295	296	297
TGLA203	298	299	300
TGLA206	301	302	303
TGLA208	304	305	306
TGLA210	307	308	309
TGLA213	310	311	312
TGLA214	313	314	315
TGLA215	316	317	318
TGLA22	319	320	321
TGLA222	322	323	324
TGLA226	325	326	327
TGLA227	328	329	330
TGLA23	331	332	333
TGLA231	334	335	336
TGLA245	337	338	339
TGLA25	340	341	342
TGLA254	343	344	345
TGLA255	346	347	348
TGLA257	349	350	351
TGLA26	352	353	354

TABLE 7 (Continued)

TGLA260	355	356	357
TGLA261	358	359	360
TGLA263	361	362	363
TGLA264	364	365	366
TGLA268	367	368	369
TGLA27	370	371	372
TGLA272	373	374	375
TGLA28	376	377	378
TGLA3	379	380	381
TGLA301	382	383	384
TGLA303	385	386	387
TGLA304	388	389	390
TGLA306	391	392	393
TGLA307	394	395	396
TGLA309	397	398	399
TGLA31	400	401	402
TGLA310	403	404	405
TGLA311	406	407	408
TGLA318	409	410	411
TGLA322	412	413	414
TGLA323	415	416	417
TGLA325	418	419	420
TGLA327	421	422	423
TGLA328	424	425	426
TGLA332	427	428	429
TGLA334	430	431	432
TGLA337	433	434	435
TGLA339	436	437	438
TGLA34	439	440	441
TGLA340	442	443	444
TGLA341	445	446	447
TGLA342	448	449	450

TABLE 7 (Continued)

TGLA345	451	452	453
TGLA346	454	455	456
TGLA35	457	458	459
TGLA351	460	461	462
TGLA353	463	464	465
TGLA354	466	467	468
TGLA357	469	470	471
TGLA36	472	473	474
TGLA37	475	476	477
TGLA377	478	479	480
TGLA378	481	482	483
TGLA380	484	485	486
TGLA381	487	488	489
TGLA382	490	491	492
TGLA387	493	494	495
TGLA39	496	497	498
TGLA394	499	500	501
TGLA4	502	503	504
TGLA40	505	506	507
TGLA400	508	509	510
TGLA414	511	512	513
TGLA415	514	515	516
TGLA417	517	518	519
TGLA419	520	521	522
TGLA420	523	524	525
TGLA421	526	527	528
TGLA423	529	530	531
TGLA424	532	533	534
TGLA427	535	536	537
TGLA429	538	539	540

TABLE 7 (Continued)

TGLA431	541	542	543
TGLA432	544	545	546
TGLA433	547	548	549
TGLA435	550	551	552
TGLA436	553	554	555
TGLA437	556	557	558
TGLA438	559	560	561
TGLA44	562	563	564
TGLA441	565	566	567
TGLA443	568	569	570
TGLA444	571	572	573
TGLA445	574	575	576
TGLA446	577	578	579
TGLA45	580	581	582
TGLA47	583	584	585
TGLA48	586	587	588
TGLA49	589	590	591
TGLA5	592	593	594
TGLA51	595	596	597
TGLA52	598	599	600
TGLA53	601	602	603
TGLA54	604	605	606
TGLA58	607	608	609
TGLA6	610	611	612
TGLA60A	613	614	615
TGLA60B	616	617	618
TGLA61	619	620	621
TGLA66A	622	623	624
TGLA67	625	626	627
TGLA68	628	629	630
TGLA69	631	632	633
TGLA70A	634	635	636
TGLA70B	637	638	639
TGLA72	640	641	642

TABLE 7 (Continued)

TGLA73	643	644	645
TGLA75	646	647	648
TGLA76	649	650	651
TGLA77	652	653	654
TGLA78	655	656	657
TGLA79	658	659	660
TGLA8	661	662	663
TGLA80	664	665	666
TGLA82	667	668	669
TGLA84	670	671	672
TGLA85	673	674	675
TGLA86	676	677	678
TGLA89	679	680	681
TGLA9	682	683	684
TGLA94	685	686	687
TGLA98	688	689	690
TGLA99	691	692	693
TGLB84	694	695	696

2. PCR-Amplification and Detection of Microsatellites

(a) Simplex amplification

Table 8 reports a preliminary list of bovine microsatellite systems that were successfully amplified in vitro, with the corresponding primer pairs. Note that pairs of primers selected by "OPTIPRIM", allow successful amplification in at least one of our standard conditions more than 80% of the time. Table 9 also gives the favoured annealing temperature (using the TECHNE MW2 heating device). The mean heterozygosity for the bovine microsatellites was estimated at ≈50%.

TABLE 8

PCR Amplified Bovine Microsatellites

Sequence Name	Up Primer Name	Up Primer ID	Down Primer Name	Down Primer ID	Annealing
AGLA13	AGLA13UP1	697	AGLA13DN1	698	55 60
AGLA206	AGLA206UP1	699	AGLA206DN1	700	55 60
AGLA209	AGLA209UP1	701	AGLA209DN1	702	55 60
AGLA215	AGLA215UP1	703	AGLA215DN1	704	60
AGLA217	AGLA217UP1	705	AGLA217DN1	706	55 60
AGLA22	AGLA22UP1	707	AGLA22DN1	708	60
AGLA226	AGLA226UP1	709	AGLA226DN1	710	55 60
AGLA234	AGLA234UP1	711	AGLA234DN1	712	55 60
AGLA254	AGLA254UP1	713	AGLA254DN1	714	55 60
AGLA255	AGLA255UP1	715	AGLA255DN1	716	55 60
AGLA258	AGLA258UP1	717	AGLA258DN1	718	55 60
AGLA260	AGLA260UP1	719	AGLA260DN1	720	55 60
AGLA269	AGLA269UP1	721	AGLA269DN1	722	55 60
AGLA284	AGLA284UP1	723	AGLA284DN1	724	55 60
AGLA285	AGLA285UP1	725	AGLA285DN1	726	55 60
AGLA29	AGLA29UP1	727	AGLA29DN1	728	55 60
AGLA291	AGLA291UP1	729	AGLA291DN1	730	55 60
AGLA293	AGLA293UP1	731	AGLA293DN1	732	55 60
AGLA8	AGLA8UP1	733	AGLA8DN1	734	55
GBFSH	GBFSHUP1	735	GBFSHDN1	736	55
GBIRBP	GBIRBPUP1	737	GBIRBPDN1	738	60
GBKCAS	GBKCASUP1	739	GBKCASDN2	740	60
MGTG1	MGTG1UP3	741	MGTG1DN1	742	55 60
MGTG13B	MGTG13BUP3	743	MGTG13BDN2	744	55 60
MGTG3	MGTG3UP1	745	MGTG3DN2	746	55 60
MGTG4B	MGTG4BUP2	747	MGTG4BDN2	748	55 60
MGTG7	MGTG7UP3	749	MGTG7DN3	750	55 60
TGLA10	TGLA10UP1	751	TGLA10DN1	752	60
TGLA111	TGLA111UP1	753	TGLA111DN1	754	60
TGLA116	TGLA116UP1	755	TGLA116DN1	756	60
TGLA117	TGLA117UP1	757	TGLA117DN1	758	60
TGLA12	TGLA12UP1	759	TGLA12DN2	760	55
TGLA122	TGLA122UP1	761	TGLA122DN1	762	60
TGLA123	TGLA123UP1	763	TGLA123DN1	764	60
TGLA124	TGLA124UP1	765	TGLA124DN1	766	60
TGLA125	TGLA125UP2	767	TGLA125DN2	768	55 60
TGLA126	TGLA126UP1	769	TGLA126DN1	770	55
TGLA127	TGLA127UP1	771	TGLA127DN1	772	55
TGLA128	TGLA128UP1	773	TGLA128DN1	774	60
TGLA130	TGLA130UP1	775	TGLA130DN1	776	55

TABLE 8 (Continued)

TGLA132	TGLA132UP1	777	TGLA132DN1	778	55
TGLA134	TGLA134UP1	779	TGLA134DN1	780	55 60
TGLA137	TGLA137UP1	781	TGLA137DN1	782	60
TGLA142	TGLA142UP1	783	TGLA142DN1	784	60
TGLA147	TGLA147UP1	785	TGLA147DN1	786	60
TGLA15	TGLA15UP2	787	TGLA15DN2	788	55 60
TGLA153	TGLA153UP2	789	TGLA153DN2	790	60
TGLA158	TGLA158UP1	791	TGLA158DN1	792	60
TGLA159	TGLA159UP1	793	TGLA159DN1	794	55 60
TGLA164	TGLA164UP1	795	TGLA164DN1	796	55 60
TGLA170	TGLA170UP1	797	TGLA170DN1	798	60
TGLA176	TGLA176UP1	799	TGLA176DN1	800	60
TGLA182	TGLA182UP1	801	TGLA182DN1	802	60
TGLA203	TGLA203UP1	803	TGLA203DN1	804	60
TGLA206	TGLA206UP1	805	TGLA206DN1	806	60
TGLA210	TGLA210UP1	807	TGLA210DN1	808	60
TGLA214	TGLA214UP1	809	TGLA214DN1	810	55 60
TGLA215	TGLA215UP1	811	TGLA215DN1	812	55 60
TGLA22	TGLA22UP1	813	TGLA22DN1	814	60
TGLA227	TGLA227UP1	815	TGLA227DN1	816	55 60
TGLA23	TGLA23UP1	817	TGLA23DN1	818	60
TGLA231	TGLA231UP1	819	TGLA231DN1	820	55 60
TGLA245	TGLA245UP1	821	TGLA245DN1	822	55 60
TGLA260	TGLA260UP1	823	TGLA260DN1	824	55 60
TGLA263	TGLA263UP1	825	TGLA263DN1	826	55 60
TGLA28	TGLA28UP3	827	TGLA28DN2	828	55 60
TGLA303	TGLA303UP1	829	TGLA303DN1	830	60
TGLA304	TGLA304UP1	831	TGLA304DN1	832	60
TGLA307	TGLA307UP1	833	TGLA307DN1	834	55 60
TGLA309	TGLA309UP1	835	TGLA309DN1	836	55 60
TGLA322	TGLA322UP1	837	TGLA322DN1	838	55 60
TGLA325	TGLA325UP1	839	TGLA325DN1	840	55 60
TGLA327	TGLA327UP1	841	TGLA327DN1	842	60
TGLA328	TGLA328UP1	843	TGLA328DN1	844	55 60
TGLA334	TGLA334UP1	845	TGLA334DN1	846	55 60
TGLA337	TGLA337UP1	847	TGLA337DN1	848	60
TGLA339	TGLA339UP1	849	TGLA339DN1	850	55
TGLA34	TGLA34UP2	851	TGLA34DN1	852	60
TGLA340	TGLA340UP1	853	TGLA340DN1	854	55 60
TGLA341	TGLA341UP1	855	TGLA341DN1	856	60
TGLA342	TGLA342UP1	857	TGLA342DN1	858	60
TGLA346	TGLA346UP1	859	TGLA346DN1	860	60
TGLA35	TGLA35UP1	861	TGLA35DN1	862	60

TABLE 8 (Continued)

TGLA351	TGLA351UP1	863	TGLA351DN1	864	55 60
TGLA353	TGLA353UP1	865	TGLA353DN1	866	60
TGLA354	TGLA354UP1	867	TGLA354DN1	868	55 60
TGLA357	TGLA357UP1	869	TGLA357DN1	870	55 60
TGLA36	TGLA36UP1	871	TGLA36DN1	872	60
TGLA37	TGLA37UP1	873	TGLA37DN1	874	60
TGLA377	TGLA377UP1	875	TGLA377DN1	876	55 60
TGLA378	TGLA378UP1	877	TGLA378DN1	878	60
TGLA382	TGLA382UP1	879	TGLA382DN1	880	60
TGLA387	TGLA387UP1	881	TGLA387DN1	882	60
TGLA40	TGLA40UP1	883	TGLA40DN1	884	60
TGLA415	TGLA415UP1	885	TGLA415DN1	886	55 60
TGLA420	TGLA420UP1	887	TGLA420DN1	888	55 60
TGLA421	TGLA421UP1	889	TGLA421DN1	890	55 60
TGLA423	TGLA423UP1	891	TGLA423DN1	892	55 60
TGLA431	TGLA431UP1	893	TGLA431DN1	894	55 60
TGLA433	TGLA433UP1	895	TGLA433DN1	896	60
TGLA435	TGLA435UP1	897	TGLA435DN1	898	55 60
TGLA44	TGLA44UP2	899	TGLA44DN1	900	55 60
TGLA441	TGLA441UP1	901	TGLA441DN1	902	60
TGLA444	TGLA444UP1	903	TGLA444DN1	904	55 60
TGLA45	TGLA45UP1	905	TGLA45DN1	906	60
TGLA47	TGLA47UP1	907	TGLA47DN1	908	55 60
TGLA48	TGLA48UP1	909	TGLA48DN1	910	55 60
TGLA49	TGLA49UP1	911	TGLA49DN2	912	55
TGLA51	TGLA51UP1	913	TGLA51DN1	914	60
TGLA52	TGLA52UP1	915	TGLA52DN1	916	55
TGLA53	TGLA53UP1	917	TGLA53DN1	918	55 60
TGLA58	TGLA58UP1	919	TGLA58DN1	920	55 60
TGLA6	TGLA6UP1	921	TGLA6DN1	922	60
TGLA60A	TGLA60AUP1	923	TGLA60ADN1	924	55
TGLA60B	TGLA60BUP1	925	TGLA60BDN1	926	55
TGLA61	TGLA61UP1	927	TGLA61DN1	928	55 60
TGLA67	TGLA67UP1	929	TGLA67DN1	930	60

TABLE 8 (Continued)

TGLA68	TGLA68UP1	931	TGLA68DN1	932	60
TGLA72	TGLA72UP1	933	TGLA72DN1	934	55 60
TGLA73	TGLA73UP1	935	TGLA73DN1	936	55
TGLA75	TGLA75UP1	937	TGLA75DN1	938	55
TGLA76	TGLA76UP1	939	TGLA76DN1	940	55
TGLA77	TGLA77UP1	941	TGLA77DN1	942	55 60
TGLA80	TGLA80UP1	943	TGLA80DN1	944	60
TGLA82	TGLA82UP1	945	TGLA82DN1	946	55
TGLA86	TGLA86UP1	947	TGLA86DN1	948	55 60
TGLA89	TGLA89UP1	949	TGLA89DN1	950	52

(b) Multiplex Amplification

To increase the speed and lower the cost of genotyping, multiplex approaches for both amplification and data capture of microsatellites are utilized. Microsatellite systems yielding products of non-overlapping size were coamplified as described above. Preliminary results show that at least four different systems can easily be coamplified in these standard conditions. The following multiplex amplifications, for instance, were shown to yield consistent, easily interpretable results:

a. GBCYP21 - TGLA10 - TGLA 44 - TGLA116
b. TGLA9 - MGTG4B - TGLA23 - TGLA35
c. MGTG3 - MGTG13B

By limiting detection to a single detection procedure (autoradiography of ^{32}P-labeled product), multiplex amplification is limited to systems yielding products of non-overlapping size. To overcome this limitation, alternative detection schemes are utilized. In particular, the use of confocal microscopy to detect products labeled with laser-excitable fluorescent molecules (such as fluoresceine, rhodamine, ...) is used. The products can then be differentiated based on the specific excitation and emission spectra of the tagged fluorescent molecules. Using this approach detection of up to at least 20 different systems should is feasable.

3. PCR-mapping of Bovine Micro-satellites Using Somatic Cell Hybrids

Results of the concordancy analysis are summarized in Table 9. Synteny groups to which microsatellite systems most likely map as deduced from concordancy analysis are underlined. Clear-cut results were obtained for MGTG13B (U19 or chromosome 15), TGLA6 (U11), TGLA9 (U27), TGLA11 (U16), TGLA22 (U26 or chromosome 26), TGLA23 (U11), TGLA36 (U27), TGLA52 (U9 or chromosome

18). Results are less discriminating for the other systems. Most likely synteny groups are, however, mentioned. In addition, we know from the literature that GBKCAS maps to U15 or chromosome 6, and GBCYP21 to U20 or chromosome 23.

TABLE 9

Concordancy Analysis for Synteny Mapping of Microsatellites

SYNT.GR.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
CHROM.			5	21	10				18				2		6	
MGTG001	90	68	70	70	40	45	50	40	69	45	70	75	65	53	40	65
MGTG04B	70	58	70	80	40	65	40	60	69	55	60	55	95	42	60	65
MGTG13A	80	47	60	70	50	55	60	30	75	55	70	75	75	63	50	65
MGTG13B	50	26	60	40	70	85	80	60	56	55	60	55	55	53	80	55
TGLA005	45	26	35	35	65	60	65	35	44	50	55	40	50	58	65	80
TGLA006	70	63	70	50	70	45	50	30	69	35	100	75	65	42	40	75
TGLA009	65	58	75	55	45	50	45	45	56	50	75	70	60	42	45	70
TGLA010	65	74	85	55	55	40	45	55	56	40	75	80	60	53	35	60
TGLA011	65	47	55	55	45	50	45	25	69	40	75	60	70	37	45	100
TGLA015	70	47	70	50	50	35	60	30	50	25	70	65	55	63	40	65
TGLA022	30	37	50	40	50	65	60	50	63	55	50	35	35	68	50	45
TGLA023	65	68	65	55	65	40	45	35	69	40	95	70	70	37	45	80
TGLA036	65	58	75	55	45	50	45	45	56	50	75	70	60	42	45	70
TGLA051	40	37	40	40	40	65	60	50	56	55	30	25	35	58	50	35
TGLA052	55	53	45	75	35	70	45	75	56	100	35	50	60	63	65	40

TABLE 9 (Continued)

Concordency Analysis for Synteny Mapping of Microsatellites

SYNT.GR.	17	18	19	20	21	22	23	24	25	26	27	28	29	X
CHROM.	8		15	23	19			14		26		24		X
MGT6001	65	80	50	55	45	45	40	35	50	40	55	40	50	65
MGT604B	95	80	50	65	45	45	60	55	33	40	55	50	60	45
MGT613A	75	80	50	55	45	55	50	55	33	40	65	50	50	55
MGT613B	55	60	100	55	45	85	50	55	17	70	55	40	70	35
TGLA005	50	55	65	50	50	70	65	60	33	55	50	55	55	50
TGLA006	65	80	60	55	25	55	50	45	83	50	75	40	60	75
TGLA009	60	65	55	70	40	60	55	40	50	45	100	65	45	50
TGLA010	60	85	55	60	30	60	45	30	67	55	70	45	45	70
TGLA011	70	75	55	70	40	60	55	40	50	45	70	55	45	60
TGLA015	55	80	50	45	35	45	30	35	33	50	55	20	40	85
TGLA022	35	60	70	55	55	85	40	55	50	100	45	40	60	45
TGLA023	70	75	55	60	30	50	55	50	83	45	70	45	65	70
TGLA036	60	65	55	70	40	60	55	40	50	45	100	65	45	50
TGLA051	35	40	60	45	65	65	30	45	33	80	25	40	50	35
TGLA052	60	45	55	70	90	70	75	80	50	55	50	75	75	10

EXAMPLE 4

CONSTRUCTION OF A PRIMARY BOVINE DNA MARKER MAP.

Bovine pedigrees for a total of approximately 200 individuals were genotyped for 150 of these markers as described. Pair-wise linkage analysis was performed using the LODSCORE program. Only lodscore values superior to 3 were considered significant. This generated a primary DNA marker map with 24 linkage groups counting two or more markers (15 assigned to specific chromosomes or synteny groups), and 68 singleton markers. Table 10 summarizes our findings. Linkage groups were assigned to specific chromosomes or synteny groups whenever that information was available.

TABLE 10

Primary Bovine DNA Marker Map

CHR./SYNT.	LG	MARKER
Chr.2	1	GMBT28
	1	GMBT47
	1	GMBT61
	1	MSBT13
	1	TGLA11
	1	TGLA44
	1	TGLA61
	1	TGLA215
	1	TGLA159
	1	TGLA58
	1	TGLA60
	1	TGLA377
	1	MGTG4B
	2	TGLA116
	2	Weaver
Chr.6		GBKCAS
Chr.8		GMBT17

TABLE 10 (Continued)

CHR./SYNT.	LG	MARKER
Chr.10		GMBT19
Chr.14		Thyroglobulin
		GMBT6
Chr.15	3	MGTG13B
	3	MSBT35
	4	GBFSH
	4	TGLA75
Chr.19		GH
		GMBT22
Chr.21	5	GMBT15
	5	GMBT16
	5	GMBT39
	5	MSBT29
	5	TGLA122
	5	TGLA337
Chr.23	6	GMBT12
	6	MSBT43
	6	MSBT70
	6	Prolactin
	6	BoLA
	6	GBCYP21
	6	AGLA291
	6	MGTG7
	6	TGLA142
	7	GMBT41
	7	MSBT6
	7	AGLA29
	7	TGLA126
	7	TGLA153
	7	TGLA214
Chr.24	8	GMBT5
	8	MSBT33

TABLE 10 (Continued)

CHR./SYNT.	LG	MARKER
Chr.26		GMBT11
		TGLA22
		TGLA51
Chr.X		GMBT27
		TGLA68
		TGLA124
		TGLA72
		TGLA54
		TGLA89
Chr.Y		DYZ1
U1	9	GMBT42
	9	MGTG13A
	9	TGLA53
U10	10	GMBT7
	10	GMBT26
	10	MSBT122
	11	TGLA52
	11	TGLA57
	11	TGLA415
U11	12	TGLA6
	12	TGLA23
U16		TGLA5
		GMBT1
	13	GMBT53
	13	TGLA206
U22		GMBT14
		GMBT8
U23		GMBT36
	14	TGLA9
	14	TGLA36
?	15	GMBT3
	15	GMBT29
	15	GMBT49

TABLE 10 (Continued)

CHR./SYNT.	LG	MARKER
?	16	GMBT21
	16	GMBT33
?	17	GMBT24
	17	MSBT15
	17	TGLA164
	17	TGLA48
	17	TGLA303
?	18	GMBT18
	18	AGLA254
?	19	AGLA226
	19	TGLA28
?	20	MGTG1
	20	TGLA245
?	21	MSBT11
	21	MSBT19
	21	TGLA227
?	22	TGLA378
	22	TGLA433
?	23	TGLA51
	23	TGLA94
?	24	TGLA54
	24	TGLA68

CHR./SYNT.: Chromosome or synteny group
LG: Linkage group

CONCLUSIONS

Samples of *E. coli* harboring clones of polymorphic bovid markers have been deposited on 17 January 1991 with the American Type Culture Collection (Rockville, Maryland) under accession numbers 68,514 and 68,515. Deposit of the clones is for the sake of completeness, but is not intended to limit the scope of the instant invention to said deposited materials. Access to said

cultures will be available during the pendency of the application to those determined by the Commissioner of Patents and Trademarks to be entitled thereto. All restrictions on availability will be removed upon grant of the application and said cultures will remain available during the life of the patent. Nonviable or destroyed cultures will be replaced in kind.

It will be appreciated that the methods and compositions of the instant invention can be incorporated in the form of a variety of embodiments, only a few of which are disclosed herein. It will be apparent to the artisan that other embodiments exist and do not depart from the spirit of the invention. Thus the described embodiments are illustrative and should not be construed as restrictive.

TABLE 11

LIST OF REFERENCES

1. NEI, M. (1987). Molecular Evolutionary Genetics, Columbia University Press, New York.

2. KAZAZIAN, H.H., et al. (1983). Evolution of Genes and Proteins (Ed. by M. Nei and R.K. Koehn), pp. 137-146. Sinauer, Sunderland, Mass.

3. GEORGES, M., et al. (1987). Animal Genetics 18: 41-54.

4. HILBERT, P., et al. (1989). Animal Genetics 20: 383-394.

5. BARKER, D., et al. (1984). Cell 36:131-138.

6. RUBIN, G.M. (1983). Mobile Genetic Elements (Ed. by J.A. Shapiro), Academic Press, pp. 329-362.

7. HAWLEY, R.G., et al. (1982). Proc. Nat. Acad. Sci. USA 79:7425-7429.

8. CANAANI, E., et al. (1983). Proc. Nat. Acad. Sci. USA 80:7118-7122.

9. YMER, S., et al. (1985). Nature 317:255-258.

10. BLATT, C., et al. (1988). EMBO J. 7:4283-4290.

11. STOCKING, C., et al. (1988). Cell 53:869-879.

12. COHEN, J.C., and VARMUS, H.E. (1979). Nature 278: 418-423.

13. LEHRMAN, M.A., et al. (1987). Cell 48:827-835.

14. KAZAZIAN, H.H., et al. (1988). Nature 332:164-166.

15. MORSE, B., et al. (1988). Nature 333:87-90.

16. LIN, C.S., et al. (1988). Cell 54:153-159.

17. WONG, Z., et al. (1990). Genomics 7:222-234.

18. JEFFREYS, A.J., et al. (1985). Nature 314:67-73.

19. JEFFREYS, A.J., and WILSON, V. (1985). Nature 316: 76-79.

20. NAKAMURA, Y., et al. (1987). Science 235:1616-1622.

21. VASSART, G. et al. (1987). Science 235:683-684.

22. SINGER, M. F. (1982). Intl. Review of Cytology 76: 67-111.

23. LATHROP, M., et al. (1988). Genomics 2:157-164.

24. LATHROP, M., et al. (1988). Genomics 3:361-366.

25. NAKAMURA, Y., et al. (1988). Genomics 3:342-346.

26. NAKAMURA, Y., et al. (1988). Genomics 2:302-309.

27. NAKAMURA, Y., et al. (1988). Genomics 3:67-71.

28. O'CONNELL, P., et al. (1987). Genomics 1:93-102.

29. O'CONNELL, P., et al. (1989). Genomics 4:12-20.

30. JULIER, C., et al. (1990). Proc. Nat. Acad. Sci. USA 87:4585-4589.

31. GEORGES, M. et al. (1990). Genomics 6:461-474.

32. WEBER, J.L. and MAY, P.E. et al. (1989). American Journal of Human Genet. 44:388-96.

33. LITT, M. and Luty, J.A. (1989). American Journal of Human Genet. 44:397-401.

34. TAUTZ et al. (1989). Nucl. Acids Res. 17:6463-71.

35. ROYLE, N.J. et al. (1988). Genomics 3:352-360.

36. NAKAMURA, Y. et al. (1987). Nucl. Acids Res. 15: 2537-2547.

37. ECONOMOU, E.P., et al. (1990). Proc. Nat. Acad. Sci. USA 87:2951-2954.

38. DOOLITTLE, W.F. (1985). Trends in Genetics 1 (3):64-65.

39. JEFFREYS, A.J., et al. (1990). Cell 60:473-485.

40. SOUTHERN, E.M. (1975). Journal of Molecular Biology 98:503-517.

41. SAIKI, R.K., et al. (1988). Science 239:487.

42. KREITMAN, M., and AGUADE, M. (1986). Proc. Nat. Acad. Sci. USA 83:3562-3566.

43. SKOLNICK, M.H., and WALLACE, R.B. (1988) Genomics 2:273-279.

44. GEORGES, M., et al. (1990). 4th World Congress on Genetics applied to Livestock Production, Edinburgh, Scotland, July 1990.

45. THEIN, S.L., and WALLACE, R.B. (1986). Human Genetic Diseases: A Practical Approach (Ed. by K.E. Davies), IRL Press.

46. SAIKI, R.K., et al. (1989). Proc. Nat. Acad. Sci. USA 86:6230-6234.

47. NEWTON, C.R., et al. (1989). Nucl. Acids Research 17(7):2503-2516.

48. WU, D.Y., et al. (1989). Proc. Nat. Acad. Sci. USA 86:2757-2760.

49. GIBBS, R.A., et al. (1989). Nucleic Acids Research 17(7):2437-2448.

50. WU, D.Y., and WALLACE, R.B. (1989). Genomics 4: 560-569.

51. MYERS, R.M., et al. (1988). Genome Analysis (Ed. by K.E. Davies). IRL Press.

52. ROBERTS, R.G., et al. (1989). Nucl. Acids Research 17(15):5961-5971.

53. ORITA, M., et al. (1989). Proc. Nat. Acad. Sci. USA 86: 2766-2770.

54. ORITA, M., et al. (1989). Genomics 5(4) :874-879.

55. CHURCH, G.M. and GILBERT, W. (1984). Proc. Nat. Acad. Sci. USA 81:1991-1995.

56. WONG, C., et al. (1987). Nature 330:384-386.

57. KIMURA, M. (1986). The Neutral Theory of Molecular Evolution, Cambridge University Press.

58. NEI, M. (1983). Evolution of Genes and Proteins (Ed. by M.Nei and R.K.Koehn). Sinauer Associates, Sunderland.

59. HILL, R.E., and HASTIE, N.D. (1987). Nature 236:96-99.

60. LATHROP, G.M. and LALOUEL, J.M. (1984). Am.J.Human Genetics 36:460-465.

61. FLINT, J., et al. (1986). Nature 321:744-750.

62. HUGHES, A.L., and NEI, M. (1988). Nature 335:167-170.

63. AJALA, F.J. (1972). Behav. Genet. 2:85-91.

64. LI, H., et al. (1988). Nature 335:414.

65. MORTON, N.E. (1955). American Journal of Human Genetics 7:277-318.

66. LATHROP, G.M., et al. (1984). Proc. Nat. Acad. Sci. USA 101:3443-3446.

67. LATHROP, G.M., et al. (1985). American Journal of Human Genetics 37:482-498.

68. LANDER, E.S. and GREEN, P. (1987). Proc. Nat. Acad. Sci. USA 84:2363-2367.

69. LICHTER, P., et al. (1990). Science 247:64-69.

70. COX, D.R., et al. (1990). Science 250:245-250.

71. WOMACK, J.E. (1987). Developmental Genetics 8:281-293.

72. WHITE, R., et al. (1985). Nature 313:101-105.

73. DONIS-KELLER, H. et al. (1987). Cell 51(2):319-337.

74. FRIES, R., et al. (1989). Animal Genet. 20:3-29.

75. LANDER, E.S., and BOTSTEIN, D. (1986). Cold Spring Harbor Symposia on Quantitative Biology 51:49-62.

76. BOTSTEIN, D., et al. (1980). Am. J. Hum. Genet. 32:314-331.

77. WHITE, R., and CASKEY, C.T. (1988). Science 240: 1483-1488.

78. PATERSON, A.H., et al. (1989). Nature 335:721-726.

79. MARTIN, B., et al. (1989). Science 243:1725-1727.

80. HANSET, R. (1982). Proceedings of the 2nd World Congress on Genetics applied to Livestock Production, Madrid, Spain, 6:439-453.

81. LANDER. E.S., and BOTSTEIN, D. (1989). Genetics 121:185-199.

82. SOLLER, M., et al. (1976). Theoret. and Applied Genetics 47:35-39.

83. WELLER, J.I., et al. (1990). in press.

84. ARNHEIM, N., et al. (1985). Proc. Nat. Acad. Sci. USA 82:6970-6974.

85. PLOTSKY, Y., et al. (1990). Proceedings of the 4th World Congress on Genetics Applied to Livestock Production, Edinburgh, Scotland, July 1990. Volume XIII, 133-136.

86. GELDERMANN, H., et al. (1985). Theoretical and Applied Genetics 70:138-146.

87. COWAN, C.W., et al. (1990). Theoret. Appl. Genet. 79:577-582.

88. HANSET, R. (1989). Annales de Medecine Veterinaire 133:465-488.

89. GROSCLAUDE, F. (1988). INRA Production Animale 1(1):5-17.

90. GIBSON, J.P., et al. (1990). Proceedings of the 4th World Congress on Genetics Applied to Livestock Production. Edinburgh, Scotland, July 1990. Volume XIV, 163-166.

91. SOLLER, M. and BECKMAN, J.S. (1982). Restriction Fragment Length Polymorphisms and Genetic Improvement, 2nd World Congress on Genetics Applied to Livestock Production, Madrid, 6:396-404.

92. HERR, C.M., et al. (1990). Proceedings of the 4th World Congress on Genetics Applied to Livestock Production, XVI:334-343.

93. MEDRANO, J.F., and AGUILAAR-CORDOVA, E. (1990). Biotechnology 8:144-146.

94. ARCHIBALD, A.L. and IMLAH, P. (1985). Animal Blood Groups and Biochemical Genetics 16:253-263.

95. MacLENNAN, D.H. (1990). Nature 343:559-561.22

96. SMITH, C., and SIMPSON, S.P. (1986). J. Animal Breeding Genet. 103:205-217.

97. WOMACK, J.E., and MOLL, Y.O (1986). J.Hered. 77: 2-7.

98. WHITE, R. and LALOUEL, J.M. (1988). Ann. Rev. Genet. 44:259-79.

99. COOPER, D.N. et al. (1985). Human Genet. 69: 201-205.

100. RUSSE, I. (1983). Bibliotheca Anatomica 24:77-92.

101. BETTERIDGE, K.J., et al. (1989). Journal of Reproduction and Fertility Suppl. 38:87-98.

102. EPPIG, J.J. (1989). J.Reprod. Fert. Suppl. 38:3-8.

103. GOULD, K.L., et al. (1990). Science 250:1573-1575.

104. HUNT, T. (1989). Nature 342:483-484.

105. SEIDEL, G.E. (1981). Science 211:351-358.

106. EPPIG, J.J. and SCHROEDER, A.C. (1989). Biol. Reprod. 41:268-276.

107. CARROLL, J., et al. (1990). Journal of Reproduction and Fertility 90:321-327.

108. MORTON, D.B., et al. (1987). Veterinary Record 121:592-593.

109. GEORGES, M., et al. (1988). J. Am. Vet. Med. Assoc. 193:9.

110. GEORGES, M., et al. (1988). Cytogenet. Cell Genet. 47:127-131.

111. FEINBERG and VOGELSTEIN (1983). Anal. Biochem. 132:6-13.

112. SEALEY, P.G., et al. (1985). Nucl. Acids Res. 13: 1805-1922.

113. ODELBERG, S.J. et al. (1989). Genomics 5:914-924.

114. THREADGILL, D.W., et al. (1990). Genomics 8:22-28.

115. FRIES, R., et al. (1986). Animal Genetics 17:287-294.

116. ISCNDA (1989). International System for Cytogenetic Nomenclature of Domestic Animals, D. DiBerardino et al., Eds., Cytogen. Cell Genet. 53:65-79 (1990).

117. PERRET, J. et al. (1990). Genomics 6:482-490.

118. EVANS, W.J., et al. (1981). Proc. Nat. Acad. Sci. USA 78:3748-3750.

119. CHAKRAVARTI, A. et al. (1984). Proc. Nat. Acad. Sci. USA 81:6085-6089.

120. SHIN, ___ et al. (19___) _____.

121. JARMAN, ___ et al. (19___) _____.

122. BELL, ___ et al. (19 ___) _____.

123. COPPRETERS, W., et al. (19___). Animal Genetics 21: 29-30.

124. JEFFREYS, A.J., et al. (1988). Nature 332:278-280.

125. JEFFREYS, A.J., et al. (1988). Nature 332:278-280.

126. HILLET et al. (1989). Anim. Genetics 20, 145-155.

127. FARBER, C.M., et al. (1989). Human Genet. 182:1-3.

128. BISHOP, D.T., et al. (1983). Am. J. Human Genet. 33:795-815.

129. SKOLNICK, M.H. and WHITE, R. (1982). Cytogenet. Cell Genet 32:58-67.

130. EWENS, W.J., et al. ()1981). Proc. Nat. Acad. Sci. USA 78:3748-3750.

131. SCHUMM, J.W. et al. (1988). Am.J.Hum.Genetics 42: 143-159.

132. COOPER, D.N., et al. (1985). Human Genetics 69: 201-205.

133. WOMACK, J.E. (1987). Development. Genet. 8:281-293.

134. RIORDAN, J.R., et al. (1989). Science 245:1066-1073.

135. SHOWS, T.B., et al. (1983). Advances in Human Genetics 12:341-452.

136. FALCONER, D.S. (1989). Introduction to Quantitative Genetics, 3rd Ed., Longman Scientific & Technical.

137. HULTMAN, T., et al. (1989). Nucl. Acids Research 17(13):4937-4946.

138. ORKIN, S.H. (1986). Cell 47:845.

139. QUESADA, M.A. et al. (1991). Biotechniques (in press).

140. HOESCHELE, I. and MEINERT, T.R. (1990). J. Dairy Sci. 73:2503.

141. GEORGES, M. and MASSEY, J.M. (1991). Theriogenology 35:151.

142. GEORGES, M. (1991). In Animal Applications of Research in Mammalian Development, Pedersen et al., Eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbour, New York, pp. 285-325.

143. GEORGES, M. et al. (1988). Nucl. Acids Res.15:7193.

144. FRIES, R. et al. (1990). In Gene Mapping: Strategies, Techniques and Applications, Schook, L.B. et al., Eds., Dekker, New York.

145. JULIER, C. et al. (1990). Proc. Nat. Acad. Sci. USA 87:4585-4589.

146. WONG, Z. et al. (1986). Nucl. Acids Res. 14:4605-4616.

147. WONG, Z. et al. (1987). Ann. Hum. Genet. 51:269-288.

148. STEELE, M.R. and GEORGES, M. (1991). Genomics 10:889-904.

149. FRIES, R. et al. (1990). Genomics 8:403-406.

150. GEORGES, M. et al. (1991). Genomics (in press).

151. LOGUE, D.N. and HARVEY, M.J.A. (1978). J. Repro. Fertil. 54:159-165.

152. BOWDEN, D.W. et al. (1989). J. Human. Genet. 44:671-678.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 307:

TTTCACAGTTGTCTAAATAAGAGAGTTATAATCACCCCACCCCCAGGTCA
TGGTCTAGTGCTCTTCTTCCAGAAAAATCCAATCTAAGCATTTGGGTGAA
GGGGGTCTGGCTGAAGACAACAGGA

(2) INFORMATION FOR SEQ ID NO: 308:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Bos taurus

(vii) IMMEDIATE SOURCE:
(B) CLONE: TGLA210 (repeat)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 308:

TGTGTGTGTGTGTGTGTGTGTGTGT

(2) INFORMATION FOR SEQ ID NO: 309:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 133 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Bos taurus

(vii) IMMEDIATE SOURCE:
(B) CLONE: TGLA210 (down)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 309:

TCTAGAGGATTCCCTGGAGAAGAGAATGGCAACCCACTCCAGTATTCTTG
CCTGGGAAATCCCATGGACAGAGGAGCCTGGTGGGTTATAGTCCATGAGG
TTGCAAAGAGTCAGACAGGACTGAATGACTAAT

(2) INFORMATION FOR SEQ ID NO: 310:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 183 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

WHAT IS CLAIMED IS:

1. A set of nucleic acid fragments that hybridize to polymorphic loci in bovids, wherein said set comprises fragments that hybridize to at least five unique loci and each fragment hybridizes to a locus comprising at least two alleles and with a heterozygosity of at least 50%.

2. The set of claim 1, wherein said polymorphic loci are selected from the group consisting of VNTR loci, multisite haplotype loci, microsatellite loci and combinations thereof.

3. The set of claim 2, wherein said polymorphic loci are VNTR loci.

4. The set of claim 3, wherein said fragments are selected from the group of VNTR markers identified in Table 1.

5. The set of claim 2, wherein said polymorphic loci are multisite haplotype loci.

6. The set of claim 5, wherein said fragments are selected from the group of multisite haplotype markers identified in Table 5.

7. The set of claim 2, wherein said polymorphic loci are microsatellite loci.

8. The set of claim 7, wherein said fragments are selected from the group of microsatellite markers identified in Table 7.

9. The set of claim 1, wherein said bovids are bovines.

10. The set of claim 9, wherein said bovines are of the species Bos taurus.

11. The set of claim 1, wherein said bovids are ovines.

12. The set of claim 1, wherein said fragments are obtained from a bovid genome.

13. The set of claim 12, wherein said bovid is a bovine.

14. The set of claim 13, wherein said bovine is of the species Bos taurus.

15. The set of claim 2, wherein said fragments are selected from the group of VNTR markers identified in Table 1, the group of multisite haplotype markers identified in Table 5, the group of microsatellite markers identified in Table 7, and combinations thereof.

16. The set of claim 7, wherein said fragments are selected from the group of microsatellite markers identified in Table 8.

17. A synteny map of microsatellite markers identified in Table 9.

18. A synteny map of VNTR markers identified in Table 4.

19. The microsatellite marker TGLA116 for the Weaver condition.

20. The microsatellite marker TGLA116 for the QTL trait of enhanced milk production in Brown Swiss cattle.

21. A set of nucleic acid fragments comprising at least one fragment selected from the group consisting of the VNTR markers identified in Table 1, the multi-site haplotype markers identified in Table 5, and the microsatellite markers identified in Table 7.

22. A process for mapping quantitative traits in bovids which comprises using the set of claim 21.

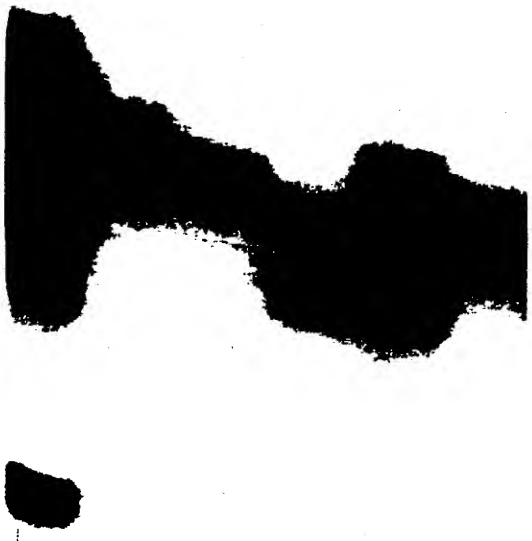
23. A process for genetic identification using the set of claim 21.

24. A process for introducing a desired gene into a bovid which comprises using the set of claim 21.

25. The process of claim 21, which further comprises the use of velogenesis.

FIG. 1

P 1 2 3 4 5



Repeat: $G(TG)_{13}$ TA $(TG)_6$ T
PCR-Product: 168bp

FIG. 2

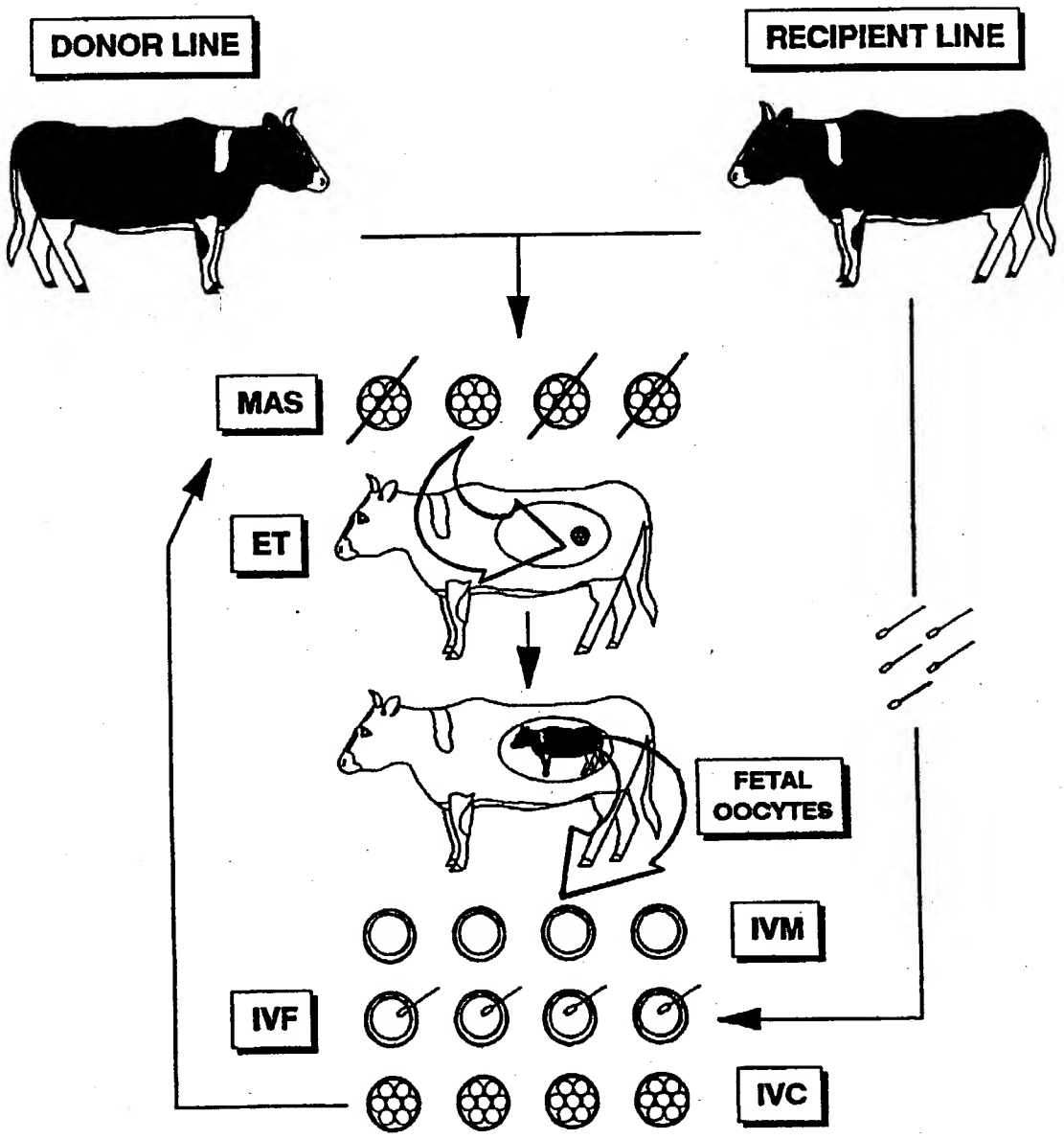


FIG. 3

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US92/00340

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[3]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC (5): C12Q 1/68, C07H 15/12, A01K 45/00
US CL : 435/6; 536/27; 119/1

II. FIELDS SEARCHED

Minimum Documentation Searched[4]

Classification System	Classification Symbols
U.S.	435/6; 536/27

Documentation Searched other than Minimum Documentation to the extent that such Documents are included in the Fields Searched[5]

CAS ONLINE, APS, MEDLINE
search terms: VNTR, MINISATELLITE, MICROSATELLITE, BOVID, CATTLE

III. DOCUMENTS CONSIDERED TO BE RELEVANT[14]

Category*	Citation of Document,[16] with indication, where appropriate, of the relevant passages[17]	Relevant to Claim No.[18]
X	WO, A, 90/15155 (Georges et al.) 13 December 1990, see all document.	1-3,5, 7-16, 21-24
X	FR, A, 2,635,116 (Georges et al.) 08 August 1988, see all document.	1-3,5, 7-16, 21-24
X	Genomics, Vol.8, issued 1990, Fries et al.,"The bovine genome contains polymorphic microsatellites", pages 403-406, see all document.	1-3,5, 7-16, 21-24
Y	Cell, Vol.18, issued November 1979, Peach et al.,"Patchwork structure of bovine satellite DNA", pages 883-893, see all document.	1-3,5, 7-16, 21-24
Y	Science, Vol.235, issued 27 March 1987, Nakamura et al.,"Variable number tendem repeat (VNTR) markers for human gene mapping",pages 1616-1622, see all document.	1-3,5, 7-16, 21-24
Y	Nature, Vol.314, issued 07 March 1985, Jeffreys et al.,"Hypervariable minisatellite regions in human DNA", pages 67-73, see all document.	1-3,5, 7-16, 21-24

* Special categories of cited documents:[15]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search[2]	Date of Mailing of this International Search Report[2]
22 APRIL 1992	07 MAY 1992
International Searching Authority[1]	**Signature of Authorized Officer[20]**
ISA/US	M. Escallon Ph.D.

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

Y	Nucleic Acids Research, Vol. 16, No.23, issued 1988, Jeffreys et al.,"Amplification of human minisatellites by polymerase chain reaction:toward DNA fingerprinting of single cells", pages 10953- 10971, see all document.	1-3,5, 7-16, 21-24
A	The Journal of Reproduction & Fertility, suppl.38, issued 1989, Betterdge et al., "Potential genetic improvement of cattle by fertilization of fetal oocytes in vitro", pages 87-98, see all document.	1-24

V. [X] OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE[1]

This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1. [X] Claim numbers 17,18,22,25, because they relate to subject matter (1) not required to be searched by this Authority, namely:

Synteny maps and gene mapping are non-statutory subject matter.

2. [X] Claim numbers 4,6,9,19,20,21-24, because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out (1), specifically:

Claims are directed to set of nucleic acid fragments supposed been disclose in a tables where only names and statistical distributions of the supposed chemical moities appear without pointout what the chemical moiety is.

3. [X] Claim numbers 25, because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI. [ ] OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING[2]

This International Searching Authority found multiple inventions in this international application as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.

2. [ ] As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3. [ ] No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

4. [ ] As all searchable claims could be searched without effort justifying an additional fee, the International Search Authority did not invite payment of any additional fee.

Remark on protest

[ ] The additional search fees were accompanied by applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category*	Citation of Document,[16] with indication, where appropriate, of the relevant passages [17]	Relevant to Claim No.[18]
Y	Genomics, Vol.6, issued 1990, Georges et al., "On the use of DNA fingerprinting for linkage studies in cattle", pages 461-474, see page 463 DNA fingerprints.	1-3,5, 7-16, 21-24
Y	The Lancet, issued 07 January 1984, Lau et al.,"A rapid screening test for antenatal sex determination",pages 14-16, see all document.	1-3,5, 7-16, 21-24
A	Nature, Vol.298, issued 19 August 1982, Streeck," A multicopy insertion sequence in the bovine genome with structural homology to the long terminal repeats of retrovirus" pages 767-769, see all document.	1-3,5, 7-16, 21-24
A	European Journal of Biochemistry, Vol.84, issued 1978, Macaya et al.,"An analysis of the bovine genome by density gradient centrifugation", pages 179-188, see all document.	1-3,5, 7-16, 21-24
A	US, A, 4,769,319 (Ellis et al.) 06 September 1988, see all document.	1-3,5, 7-16, 21-24
A	US, A, 4,960,690 (Ellis et al.) 02 October 1990, see all document.	1-3,5, 7-16, 21-24
A	US, A, 5,032,501 (Milner) 16 July 1991, see all document.	1-3,5, 7-16, 21-24
A	Nature, volume 251, issued 27 September 1974, Botchon, "Bovine satellite I DNA consists of repetitive units 1,400 base pairs in length," pages 288-292, see entire document.	1-3,5,7-16,21-24